Summary of Revisions

Thank-you for your reviews, which have helped us improve the manuscript. We have incorporated responses to all review questions and suggestions in the revised manuscript.

In all three reviews, several questions occurred about two topics: measurement limitations and the statistical approach of differencing time-synchronized concentrations. We have addressed these questions at appropriate places in the manuscript. We also provided more context and explanation by expanding the brief overview description of the study and study objectives that appears at the end of the introduction (lines 78 – 100), where we added the following paragraph.

"The mobile sampling discussed here and in Apte et al. (2017) is limited to weekdays between ~9 a.m. and 5 p.m. Sampling is necessarily conducted along roads and streets. Depending on the number of repeated driving segments, vehicles sample different road segments on different days or at different times of day. These limitations are important considerations for studies whose goal is to develop pollutant maps that represent long-term concentration averages, and which are intended to correctly characterize spatial variations at a desired spatial scale. Our objectives are different, however. The principal objectives of our study are to examine the capabilities of research instruments when placed in stationary and moving vehicles, to compare our measurements with those obtained from stationary air quality monitors, to evaluate driving and sampling strategies, and to develop statistical methods that account for sampling limitations. Limitations that are specific to our study are that (1) it was conducted as a series of geographically separated sampling campaigns between May 2016 and September 2017, generally lacking the number of repeated driving routes previously used to generate pollution maps (Apte et al., 2017; Messier et al., 2018), and (2) no collection of driving routes completely covered any specific geographical domain (e.g., San Francisco or specific neighborhoods therein). The results presented here therefore focus on measurement and methodological questions that can be addressed with data available from the individual sampling campaigns. A set of research questions was developed initially and was then used to design the individual sampling campaigns. In analyzing the results, a need arose to distinguish between temporal variability (due, e.g., to sampling different places at different times) and spatial variability. Statistical methods were therefore developed to characterize spatial heterogeneity within and between neighborhoods by utilizing time-synchronized differences in the pollutant concentrations that were measured by different vehicles. Due to limited repeated sampling of individual road segments, our estimates of spatial heterogeneity do not in themselves identify specific spatial coordinates of long-term high and low pollutant concentrations. However, areas with high spatial heterogeneity indicate where more intense future sampling would be warranted. Additional statistical methods were developed to demonstrate the use of short-term campaign measurements to characterize intermediate-scale (1 km) spatial variations of pollutant concentrations."

Responses to Anonymous Referee #1

Thank-you for your helpful questions and suggestions. We summarize our responses here. We added these responses at appropriate places in the revised manuscript or in the supplement.

1. Some information on the cars should be included, not just referenced in another paper. Were the cars' engines running while parked (e.g., in the garage, near stationary monitors, etc.)? What was used to power the instruments?

The instruments were switched from vehicle to line power in the parking garage or parking lot.

2. It is great that the inlets were designed to minimize self-sampling, but were additional steps taken during post-processing to remove potential periods of self sampling, or of sampling the Google car in front?

The cars generally followed different routes, such as those illustrated in Figure 7 or in Figures S3, S6 – S8, and S10 – S15. Therefore, sampling the exhaust of a partner car was seldom an issue. When the cars traveled a route segment together (e.g., Figure S15), they could not travel side-by-side because that would block the flow of traffic. The drivers instead traveled "caravan style", keeping each other in sight but not following immediately one behind the other.

3. It would be good to document the limitations of the study (e.g., no overnight monitoring on roads or in early morning when the boundary layer is likely at its lowest).

Please see the new paragraph above and following line 77 in the revision.

4. Table 1 - was winter included in the San Joaquin Valley measurements (Nov '16-Apr '17) or was it just fall and spring?

Sampling was conducted in the San Joaquin Valley between November 2016 and April 2017. We did not attempt to analyze the full set of measurements because another manuscript will likely be needed to fully describe the intracity, intercity, and urban-rural differences encountered in this geographically large domain. No single area was sampled throughout the entire period.

5. Section 2.2 - were the cars parked on the roof of the parking structures or on a lower level? Depending, this could explain why GPS uncertainties were not comparable to manufacturer specs at times. Tall buildings nearby (if present) would also impact GPS performance.

In San Francisco, the cars were parked within a parking structure. In Los Angeles, the cars were parked in a small (~30 car) open parking lot. We corrected this statement (e.g., at old lines 109, 113, 165) in the revision. We do not have an explanation for the observed variations in the GPS location uncertainties, but report them for completeness.

6. Lines 286-289: the closely-spaced moving vehicle condition makes it highly likely that the following car is measuring exhaust emissions from the lead car. This point should be mentioned in the manuscript. Did you try to correct for this? Why not drive side-by-side (road permitting)?

Please see our response to #2.

7. Could you please provide a list or table in the SI with the manufacturer and model of all instruments used in the study along with response time and measurement frequency. Even if this information was referenced in another paper, it should be reported here.

We added this information to the supplement (new Table S1). The instrumental methods, resolution, ranges, and response times are from the Lunden and LaFranchi (2017) citation.

8. Did you sync all instruments to the same time standard before measurements? Did you check instrument times at the completion of each day's measurement to quantify time drift? At measurement frequency of 1Hz, time drift can have a major impact on data comparison. These details should be included in the methods section.

We clarified this point at old line 137. The on-board computers are synchronized throughout the day using network time protocol, which synchronizes computers to Coordinated Universal Time (UTC) with accuracies on the order of milliseconds. This approach was necessary to ensure that the 1 Hz measurements did not drift in time.

9. By comparing 1-min averages, which I understand is important in order to maintain higher spatial resolution, how are you able to separate out the spatial trends due to differences in regional concentrations as opposed to differences in measurements due to some very localized conditions (e.g., driving behind a truck for a short period of time with one Google car but not the other over the same time period)? Would some other comparisons be more appropriate, such as a 60-second moving 5th percentile, or something comparable, to smooth out hyper-local concentrations?

Several different comparisons were made at 1-minute or coarser resolution. For the Los Angeles car-to-car comparisons, we examined both bin-average FAMD and variability within bin averages. Random differences between vehicles, such as short, intermittent exposures of one car or the other car to a high emitter, are averaged out in the FAMD statistic. In contrast, systematic car-to-car differences yield higher FAMD values. We identified some specific geographical patterns associated with higher FAMD (old lines 352 – 355). As noted, the approach developed for studying the San Francisco data could also be applied to the Los Angeles data for a more comprehensive analysis.

For the San Francisco data, we aggregated 1-minute differences to a 1-km spatial scale (old lines 402 – 410). Large mean differences were plotted in Figure 6 only if they were statistically different from zero (i.e., the interval of the mean difference ± 2 standard errors of the mean did not cover zero) so that atypical car-to-car comparisons did not artificially create apparent spatial patterns. The rationale is that the standard errors of the 1-km averages would be large if one or more paired differences was very large; this would indicate the occurrence of an unusual condition.

10. I am struggling to understand why plotting the measurements against distance between either the cars or between car and stationary monitor is the best way to present the data. Had the cars been driving different routes than the ones presented, the plots would be completely different? The distance between

the cars is not driving the differences observed, it is the difference in the environments of the two cars at any given time. For example, the cars could both be in heavy traffic at 50 km away from each other (thus mean differences in concentrations are low), then at 75 km distance one car is still in heavy traffic while the other is in a quiet neighborhood away from highways (thus mean differences in concentrations are high). For example, Figures S16 and S17 are interesting, but it would be more informative to provide information on where each of the cars are (e.g., land use, traffic conditions, major roadway, etc.) when FAMD is higher or lower irregardless of the distance between the cars. Are all points where the cars are X distance away from each other aggregated together even if the positions were discontinuous? If so, I do not know how one could interpret this plot.

We added text in the introduction and in preceding the definitions of the statistical metrics to explain how we can interpret the plots of differences versus intervehicle distance to characterize the spatial scales of pollutant heterogeneity. This issue clearly affects the data from the San Joaquin Valley, where the cars were separated by larger distances than in Los Angeles or San Francisco. As noted by the reviewer and as shown in Figures S15 and S16, for some species, differences in environments can drive car-to-car differences when the vehicles are separated by more than a few kilometers. In contrast, Figure S17 shows the expected regional character of ozone with FAMD values < 0.2 at all intervehicle distances < 50 km. We conclude that smaller FAMD values indicate greater spatial homogeneity; larger FAMD values require further study beyond the plots of FAMD vs. distance. High FAMD values indicate where further study would be informative. As one example, Tables S2 – S5 summarize car-to-car comparisons that are stratified by sampled areas. We expect that more complete analyses of the San Joaquin Valley data will be quite informative but will require another manuscript to fully explore.

11. Section 3.6 (lines 476-479): An FAMD of 0.5 seems high to conclude that a reference monitor is representative of a neighborhood scale area.

We revised this statement to note that the majority of the NO_2 FAMD values were less than 0.2 at car-monitor distances of 0.5 – 4 km. We noted the higher FAMD for NO. EPA defines this monitor as neighborhood scale for O_3 and NO_2 but not NO.

Responses to Anonymous Referee #2

Thank-you for your constructive review. We summarize our responses to your questions and suggestions here. We have added these responses at appropriate places in the revised manuscript or the supplement.

1. There is no discussion or comparison between Aclima instrumentation and capabilities to other sensors in the market (e.g. Purple Air), including technical and accuracy information. Have the authors done any comparison studies at similar times and locations to demonstrate Aclima outperforming other sensors?

Our study used only measurements from research-grade instruments (lines 78 – 79) and we added instrument specifications as new Table S1. Aclima has conducted sampling efforts using sensors during the past year. When the sensor data are analyzed, they can be compared to other sensor-based studies.

For this manuscript, we focused on comparisons to EPA-approved equipment at stationary sites in Los Angeles.

2. This analysis provides information on mobile air quality monitoring in a certain environment. The measurements represent air quality in urban locations near roads and that covers certain points/line measurements yet does not create a continuous air quality map.

Please see the new paragraph above and in the introduction at old line 77.

3. PN is measured by the Aclima platform for different size bins. It is not clear how this measurement is evaluated, as the EPA monitors particulate matter mass concentration?

PN measurements were evaluated as described in old lines 148 – 153. We were not able to do an "apples-to-apples" field comparison to EPA monitors, as the reviewer notes. Nor could we do laboratory zero and span checks, as was done for the gas instrumentation.

4. It is not clear why the distance between cars is important in the discussion.

We added text to discuss the utility and limitations of intervehicle variability versus distance. Please see also the new paragraph in the introduction (line 79). Because our study was conducted as a series of short-term campaigns in several widely separated geographical areas, we did not attempt to develop pollutant maps that represent long-term concentration averages and which could be used to characterize spatial variations. Our study was conducted as a series of geographically separated sampling campaigns between May 2016 and September 2017, generally lacking the number of repeated driving routes needed to generate stable, long-term pollution maps. Instead, we used statistical metrics, such as FAMD, to characterize the spatial heterogeneity of pollutant concentrations. Because vehicles sample different road segments on different days and at different times of day, we compiled time-synchronized differences between the concentrations measured by two cars to remove the confounding effects of day-to-day and diurnal variability. Random differences between vehicles, such as short, intermittent exposures of one car or the other car to a high emitter, are averaged out in the FAMD statistic. In contrast, systematic car-to-car differences yield higher FAMD values. Systematic differences could occur if the instrumentation in one car was biased relative to the other car. After eliminating that source of systemic car-to-car difference through the side-by-side sampling comparisons, we can conclude that larger FAMD values (e.g., > 0.20 or 20%) represent spatial heterogeneity, e.g., due to the two cars sampling different neighborhoods (as indicated in Figure 6a or in Figures S6 and S10). Considering the relationship between FAMD and distance on a small (1 – 10) number of days provides a measure of the spatial scales over which concentrations changed by more than a specified amount (e.g., 20%). This is a useful metric for evaluating the spatial scale of representativeness of stationary monitors, for example. The relationship between FAMD and distance does not, of course, indicate which neighborhoods experienced higher pollutant concentrations. For that purpose, we developed the visualization shown in Figure 6.

5. All the measurements have been done for periods of several weeks and there is no ‘long-term’ monitoring campaign presented (e.g >1 year) that captures, for example, seasonality. This limitation of measurements period should be addressed in the discussion.

Please see the new paragraph in the introduction (line 79) and our response to #5.

6. A description of the climatology at the different measurement locations is missing (e.g. temp, RH, and wind profiles, built area, type of road, no. of cars etc.). That can help understand some of the results.

By focusing on time-synchronous car-to-car measurement differences, we ensure that both vehicles are experiencing the same meteorological conditions. The figures and photos (Figures 6 and 7; Figures S3 – S4, S6 – S10) provide an indication of road density, built area, and proximity of driving routes to freeways. Population data for cities in the San Joaquin Valley are provided in lines 437 – 440 to complement Figure 7. Figures S11 – S15 indicate when the driving routes were in San Joaquin Valley cities and when they were on freeways. We added text to better highlight how this information was used, or can be used, to help interpret the results.

7. The authors should do a better job in stating the limitations of the Aclima platform in this study set and in general.

Please see the new paragraph in the introduction (line 79).

8. Did the authors consider validating their results with continuous modeled data (CMAQ)? Or satellite data?

Because we focused on interpreting the results of a series of short-term campaigns, we did not compile pollution maps. Comparison of pollution maps generated from stable, long-term data to satellite data or modeling predictions could indeed provide complementary corroborating results. For such a comparison, one challenge would be the incommensurability of the fine-scale mobile data and the coarser spatial scales of gridded modeling output or satellite imagery. The mobile data would need to be aggregated to the coarser scales for the comparison. Presumably, if results on consistent spatial scales were reasonably consistent, it would then be valuable to compare mobile monitoring maps generated from spatially-aggregated and -disaggregated data to better understand what is gained by the high-resolution mobile sampling.

Responses to Anonymous Referee #3

Thank-you for your thorough review. We summarize our responses to your questions and suggestions here. We added these responses at appropriate places in the revised manuscript or in the supplement. Please note the new paragraph that we propose to add to the introduction.

We organize our responses by the line numbers noted in the review under the two review categories

Specific Comments:

Line 17: The time frame is misleading. Should indicate a few intensive (i.e. week to month long) campaigns were performed between May 2016 and September 2017.

Revised to: On-road measurements of air quality were made during a series of sampling campaigns between May 2016 and September 2017 at high…

Lines 20-21: Lifetime of NO2 is hours and O3 is >20 days in the troposphere. Observing the diurnal cycle and weekday weekend trends may be more appropriate than looking at a fortnight.

Our focus was on characterizing spatial rather than temporal variations. As noted in the proposed new paragraph, the cars drive weekdays between ~9 a.m. and 5 p.m. The sampling regime therefore does not permit weekday/weekend comparisons nor does it lend itself to fully characterizing diurnal cycles.

Line 22: In-situ instrument or research-grade instruments. I'm sure they mean their instrument package.

Revised to: … research instruments located within stationary vehicles…

Line 31: Percentages of what? Concentration deltas?

Revised (here and at line 424) to: 1-km scale differences in NO_2 and O_3 concentrations up to 117% and 46%, respectively, of mean values

Line 75: Concentration decay rate from a point source will be highly variable and based on several meteorological parameters.

Revised to: …could help establish concentration decay rates of mobile emissions with distance…

Line: 109-110: What are the limitations of overnight calibration when cars are parked next to each other?

Lines 109 – 110 simply state that we used the QD1 data because QD1 data sets included the time periods when the cars were parked next to each other, whereas such times had been filtered out of the QD2 data set. The instruments were switched from vehicle to line power when parked.

Lines 106,109/110, 113: The first lines seem to imply the mobile platform intercomparison was made overnight, 113 implies it may have been only a short period (5 min, 30 min), the SI from their Apte et al. indicates it was several hours overnight. Not sure about an intercomparison in a parking garage either, especially if it was during a time when vehicles were entering or leaving (cold starts vs operating temp emissions).

Revised to: … with additional sampling occurring while vehicles were parked in San Francisco and Los Angeles before (~ 6 – 9 a.m.) and after (~5 – 10 p.m.) driving periods.

The vehicles are parked away from traffic in a designated area in the San Francisco parking garage. They were parked overnight in a small (~30 car) lot in Los Angeles.

Line 119: Was the audit the same as is done with FRM/FEM monitors via the National Performance Audit Program (NPAP)?

Please see Section 3.1.

Line 122: In table 4, there needs to be an explanation of the scales and how EPA initially established each sitting.

We added citation to Appendix D to Part 58 - Network Design Criteria for Ambient Air Quality Monitoring (https://www.law.cornell.edu/cfr/text/40/appendix-D_to_part_58). The EPA scales are defined in Footnote 1 of Table 4.

Line 125: Suggested to add sentence or phrase to cover why the other stations were not used.

Revised lines 119 – 125 to: During the Los Angeles sampling, the South Coast Air Quality Monitoring District (SCAQMD) conducted calibration checks when the sampling vehicles were parked adjacent to stationary air quality monitoring sites (Table 3). The SCAQMD also prepared 1-minute resolution data files for measurements made at these and other stationary air quality monitoring sites (Table 4; see also location map, Figure S1). Data from one of the dates and locations (LAXH, September 20, 2016) were suitable for collocated comparison with mobile measurements (Table 3). The stationary-monitor data from W710 consisted only of 1-hour resolution $PM_{2.5}$ mass (Table 4), which was not measured by the mobile platforms, and no data were provided for the Santa Clarita site (Tables 3).

Line 137: Do the monitoring stations use the Network Time Protocol? If not, address discrepancies this may cause in measurement comparisons.

We used time series plots, such as Figure S5, to confirm the alignment of mobile and station minima and maxima. The results indicate that any temporal discrepancies are less than the 1-minute averaging times.

Line 164: What is the impact of wind and GPS location uncertainties on data collected while stationary?

When cars were parked adjacent to each other, we do not expect GPS location uncertainties or variations in wind speed or direction to impact the side-by-side comparisons.

Sec. 2.3: Are there multiple BC and CH4 instruments or just one that was moved between cars? I'm assuming this was done because of inlet restrictions.

One car was equipped with a BC instrument and one car was equipped with a CH4 instrument. There are four cars, though. Two cars were used by Apte et al. (2017) in their study. Since all vehicles parked in the same San Francisco garage, there were two BC and two CH4 instruments available for the side-by-side parked comparisons (line 180, Tables 5 and 6).

Line 173: Where are the reported variabilities of the paired differences shown?

These are reported in "Results and Discussion," rather than "Methods" (Section 3.2, Table 9).

Sec. 2.3: Was CARB contacted to ensure BC and CH4 observations were not present at sites? Some EPA sites have but don't advertise these observations.

For the field comparisons, we worked with South Coast Air Quality Management District staff, who operate the air quality monitors and are familiar with all measurements made.

Lines: 247- 249: Comparing different regions during different time periods without a detailed study of the meteorology is misleading if talking about local or neighborhood scales.

Here are the climatological winds near San Joaquin Valley for March and November using data between 1973-2019 at Buchanan Field Airport in Concord, CA. Figure 1 shows the month of March may be experiencing inflow from the Chevron processing plant in Richmond and dust (Coarse mode, not reported) from Dutra Materials quarry in McNears Beach, while to a much lesser extent in November (Figure 2). Since the data is presented as mean concentrations during the sampling periods, I'd bet the baseline PN concentrations are different for the two months.

Lines 247 – 260 provide a summary overview of the measurements. These are useful but require caveats for various reasons such as those indicated by the referee. As we stated at line 250, "Although differences in the PN distributions possibly reflect spatial variability, they more likely reflect seasonal variations in PM composition" and at lines 261 – 262, "As with PN, these average concentrations likely vary due to time of year, location relative to source emissions, and chemical processing." For clarity, we revised these lines to: "Although these differences in the PN size distributions possibly reflect regional-scale spatial variability, no simple comparison among regions is possible due to sampling them during different seasons. The regional differences could reflect seasonal variations in PM composition: the observed variations in PN distributions are consistent with past studies that indicate the importance of PM nitrate (NO3) found in larger (> 0.5 μm) size fractions primarily as ammonium nitrate in California during cooler months (e.g., Herner et al., 2005), which could lead the observance of different size distributions in the different regions". For later analyses, we focused on time-synchronous differences between measurements made by two vehicles, not averages over short-term campaigns.

Lines 251-252: The deployed optical particle counter provided five size ranges why report only the smallest, then reference a paper regarding a measured size bin that was not reported in the paper?

Please see previous response.

Lines 246-262: I'm not sure this section is representative and should be included here and should likely be absorbed by the following sections.

As noted previously, this paragraph provides a summary overview of the measurements but is not the basis for our analyses of spatial variability.

Line 265: Typo. … vehicles drove in the Los Angeles

Revised by removing "the"

Line 270: Why was the mean relative difference between the two calibrations so high? An absolute difference of 5% NPAP would require corrective actions. The calibration gases and flow meters used should be traceable to NIST for re-evaluation.

We do not see large (>5%) differences between the internal and external calibration checks in Table 8. The invalidating limits for the South Coast Air Quality Management District's weekly calibration checks are 7% for O_3 and 10% for CO, SO_2, and NO_x, warning limits are 5% for O_3 and 7% for CO, SO_2, and NO_x (Table 2.4, https://ww3.arb.ca.gov/aaqm/qa/pqao/repository/district_sops/south_coast/quality_assurance/qapp_criteria_pollutants.pdf).

Sec. 3.2: When comparing inter-vehicle observations were the vehicles traveling the same route (i.e. following each other) or just driving the same neighborhoods and passing by each other?

The cars generally followed different routes, such as those illustrated in Figure 7 or in Figures S3, S6 – S8, and S10 – S15. When the cars traveled a route segment together (e.g., Figure S15), they traveled "caravan style", keeping each other in sight but not following immediately one behind the other.

Line 294: Were the vehicles were running during the LAXH comparison or were the instruments moved to shelter power and the vehicle engines shutoff?

The instruments were switched from vehicle to line power in the parking garage or parking lot but this option was not logistically practical for the one-day comparison at LAXH.

Sec. 3.3: Last sentence of section, CH4 emissions from vehicles is extremely small (something like <0.2% of anthropogenic emissions) and the lifetime of NO very short. This statement needs a citation, or it needs to be removed.

We added citation: Nam et al., ES&T, 2004 "We recommend the use of an average emission factor for the U.S. on-road vehicle fleet of (g of CH_4/g of CO2)) = $(15\pm 4)x10^{-5}$ and estimate that the global vehicle fleet emits 0.45 ± 0.12 Tg of CH_4 yr^{-1} (0.34 ± 0.09 Tg of C yr-1), which represents <0.2% of anthropogenic CH_4 emissions." https://pubs.acs.org/doi/10.1021/es034837g.

We agree that NO has a short residence time compared to CH_4. However, a correlation between NO and CH4 will be observed when sampling fresh automotive exhaust emissions. The revised sentence reads: "CH_4 is reported in motor-vehicle emissions (Nam et a., 2004), so a correlation between NO and CH_4 will usually be observed when sampling fresh automotive exhaust emissions; all NO values correlated with Coltrane CH_4 concentrations (r^2 = 0.84 to 0.87; Flora did not report CH_4)."

Sec 3.4: This section will have a very large dependence on meteorological parameters.

We agree that meteorology impacts the concentrations measured at two distant points, as do emission sources and chemical and physical processing. The last sentence of the first paragraph in section 3.4 states "The intent of the analyses in this section is to help elucidate the spatial scales over which stationary-monitor and mobile-platform data represent ambient concentrations." The analyses utilize time-synchronous differences so that each vehicle is experiencing the same meteorological conditions.

Sec. 3.4.1: The air masses the vehicles are sampling are potentially different. An intervehicle comparison could be made in time and latitude. As it is, the comparisons are meaningless because we know the location of any vehicle at any given time and one may be sampling south of the Santa Ana Freeway and the other sampling all three major N-S freeways in the area. The attached Figure 3 shows winds are between 9am and 5pm averaged over Aug 3-16. See comment above, section 3.4.

Because vehicles sample different road segments on different days and at different times of day, we compiled time-synchronous differences between the concentrations measured by two cars to remove the confounding effects of day-to-day and diurnal variability. Random differences between vehicles, such as short, intermittent exposures of one car or the other car to a high emitter or variations in wind directions, are averaged out in the FAMD statistic. In contrast, systematic car-to-car differences yield higher FAMD values. Systematic differences could occur if the instrumentation in one car was biased relative to the other car. After eliminating that source of systemic car-to-car difference through the side-by-side sampling comparisons, we can conclude that larger FAMD values (e.g., > 0.20 or 20%) represent spatial heterogeneity, e.g., due to the two cars sampling different neighborhoods (as indicated in Figure 6a or in Figures S6 and S10). FAMD is a useful metric for evaluating the spatial scale of representativeness of stationary monitors, for example. The relationship between FAMD and distance does not, of course, indicate which neighborhoods experienced higher pollutant concentrations. For that purpose, we examined maps, such as shown in Figures S6 and S10, and photos such as those provided by the referee; we also developed the visualization shown in Figure 6. Please note also our interpretation at lines 352 – 355.

Line 362: Driving near as in right past along Dowlen Dr or within n meters? Wilshire Blvd is ~200m as is Federal Ave.

Revised to: Driving routes were near (<0.2 to 5 km) the west Los Angeles stationary monitor (WSLA, Table 4) on four of the 14 days between September 12 and 30 (including areas shown in Figure S8 for September 13 and 19; similar routes were driven on September 26 and 29).

Line 392: What grid is used?

Nearest kilometer as calculated by conversion of latitude and longitude to UTM coordinates. Revised to:" One-minute averages were next averaged spatially to the nearest kilometer (based on conversion of latitude and longitude to Universal Transverse Mercator [UTM] coordinates) separately for each car (Figure 6b)…"

Fig. 6: Needs legend, different colors for positive and negative intervehicle differences and FMD differences not red/blue, which were used to identify specific vehicles in the same figure.

Figure revised using different colors. Subpanels are defined in the caption.

Line 424: Enhancements based on what? FAMD is comparing observations at the same time, is the enhancement based on location as stated in the paragraph before or between May 1-12?

Revised to: 1-km scale differences in NO_2 *and* O_3 *concentrations up to 117% and 46%, respectively, of mean values*

Line 440: Routes for November 16th, 2016 are not in SI but referenced in text.

Revised to: The initial drives occurred November 16 – 23, 2016 (Figure 7, November 16; see also example of drives on other days in Figures S11 – S15).

Include Line 457: Are traffic count data available?

We did not use traffic count data in our analyses but they are available. Revised to: “High traffic volumes (~50,000 – 150,000 vehicles per day, annual average peak volumes) are typical of Highway 99 (https://dot.ca.gov/programs/traffic-operations/census/traffic-volumes, last access April 15, 2020),…”

Line 459: Enhancements compared to what, background?

Lines 459 – 460 state “…enhancements of pollutant concentrations ***in northern San Joaquin Valley cities*** *over* ***concentrations occurring in surrounding areas”***

Line 529: Enhancements based on what?

Revised to: 1-km scale differences in NO_2 *and* O_3 *concentrations up to 117% and 46%, respectively, of mean values*

General: Overall distance bins should be the same for all missions. Seems like all the analysis times were weekday (do Google Street View vehicles drive on weekends)?

The spatial scales of the sampling routes differed among the missions, so the distance bins also differ. As noted at line 112, measurements were made between ~ 9 a.m. and 5 p.m. on weekdays. We identify this as a limitation in the new paragraph in the introduction.”

TECHNICAL CORRECTIONS: ——————————————

Line 35: Suggested to add spatial variability context for pollutants to introduction as this has implications on reported uncertainties. Seems this is provided starting at about line 48 of the intro.

Lines 35 – 60 provide this context. It isn't evident that reordering sentences would improve clarity.

Line 155: LOD is defined in Table 5 subtext, but not in text. Consider defining in main text.

Revised old line 153 to: We calculate BC limit of detection (LOD) (see footnote 2, Table 5) using data reported…

Lines 172-174: Suggested to remove ‘merge’ detail, as it seems superfluous to the reader, and combine the two sentences into one focusing on temporally coincident pairing.

Revised to: Data files were merged by 1-s or 1-minute resolution times and were then used to determine time-matched paired differences, which were evaluated as functions of ambient concentration, intervehicle distance, and vehicle speed.

Lines 185, 190, 195, 200: ‘Car B Difference’ could be misleading. It is suggested to move the word ‘Difference’ to after the word ‘Mean’ (i.e., Mean Difference) or use wording such as ‘Mean [Absolute] Difference between Car A and B’ in the numerator.

These changes were made.

Line 206: Z is not defined.

Z is simply an example variable, not a measurement. Lines 205 – 209 were replaced with new citations and “Z” no longer is used.

Line 216: MD already defined in line 185.

Not meant to redefine MD, just restating for clarity; revised to remove “(MD)”

Lines 211 and 222: Consistency in section references.

Now capitalized in both locations.

Mobile-Platform Measurement of Air Pollutant Concentrations in California: Performance Assessment, Statistical Methods for Evaluating Spatial Variations, and Spatial Representativeness

Paul A. Solomon[1], Dena Vallano[2], Melissa Lunden[3], Brian LaFranchi[3], Charles L. Blanchard[4], Stephanie L. Shaw[5]

[1]Independent consultant, Henderson, Nevada, 89052, USA, formerly with U.S. Environmental Protection Agency, Office of Research and Development, Las Vegas, NV, 89119, USA
[2]U.S. Environmental Protection Agency, Region 9, Air and Radiation Division, 75 Hawthorne St, San Francisco, CA, 94105, USA
[3]Aclima, Inc., 10 Lombard St, Suite 200, San Francisco, CA, 94111, USA
[4]Envair, 526 Cornell Avenue, Albany, CA, 94706, USA
[5]Electric Power Research Institute, 3420 Hillview Ave, Palo Alto, CA, 94304, USA

Correspondence to: Charles L. Blanchard (charles.ld.blanchard@gmail.com)

Abstract. Mobile platform measurements provide new opportunities for characterizing spatial variations of air pollution within urban areas, identifying emission sources, and enhancing knowledge of atmospheric processes. The Aclima, Inc. mobile measurement and data acquisition platform was used to equip four Google Street View cars with research-grade instruments, two of which were available for the duration of this study. On-road measurements of air quality were made during a series of sampling campaigns between May 2016 and September 2017 at high (i.e., 1-second [s]) temporal and spatial resolution at several California locations: Los Angeles, San Francisco, and the northern San Joaquin Valley (including non-urban roads and the cities of Tracy, Stockton, Manteca, Merced, Modesto, and Turlock). The results demonstrate that the approach is effective for quantifying spatial variations of air pollutant concentrations over measurement periods as short as two weeks. Measurement accuracy and precision are evaluated using results of weekly performance checks and periodic audits conducted through the sampler inlets, which show that research instruments located within stationary vehicles are capable of reliably measuring nitric oxide (NO), nitrogen dioxide (NO_2), ozone (O_3), methane (CH_4) black carbon (BC), and particle number (PN) concentration with bias and precision ranging from <10 % for gases to <25 % for BC and PN at 1-s time resolution. The quality of the mobile measurements in the ambient environment is examined by comparisons with data from an adjacent (< 9 m) stationary regulatory air quality monitoring site and by paired collocated vehicle comparisons, both stationary and driving. The mobile measurements indicate that U.S. EPA classifications of two Los Angeles stationary regulatory monitors' scales of representation are appropriate. Paired time-synchronous mobile measurements are used to characterize the spatial scales of concentration variations when vehicles were separated by <1 to 10 kilometers (km). A data analysis approach is developed to characterize spatial variations while limiting the confounding influence of diurnal variability. The approach is illustrated using data from San Francisco, revealing 1-km scale ~~enhancements~~ differences in mean NO_2 and O_3 concentrations up to 117 % and 46 %, respectively, of mean values during a two-week sampling period. In San Francisco and Los Angeles, spatial variations up to factors of 6 to 8 occur at sampling scales of 100 – 300m, corresponding to 1-minute averages.

1 Introduction

In 2017, air pollution was responsible for nearly 5 million premature deaths worldwide, a 5.8 % increase from 2007 (Stanaway et al., 2018). Model projections indicate a possible doubling of premature mortality due to air pollution between 2010 and 2050 (Lelieveld et al., 2015). Multiple studies associate exposure to nitrogen dioxide (NO_2), particulate matter, carbon monoxide (CO), ozone (O_3), and sulfur dioxide (SO_2) with adverse health effects (Stieb et al., 2002; U.S. EPA, 2008; 2010a; 2010b; 2014; 2018; WHO, 2006).

Over the last 45 years, the public has relied on air quality information from stationary regulatory monitoring sites that are sparsely located throughout the U.S. With the advent of air quality monitoring equipment that can be placed across a range of locations using various sampling platforms (personal, stationary, and mobile), a greater spatial and temporal understanding of air quality can be obtained. With this information, members of the public can potentially reduce their health risks from air pollution. Improved understanding of spatial variations in air pollutant exposure is expected to yield increasingly accurate estimates of the health effects of air pollution and is an important step in effectively reducing human exposure, acute and chronic health impacts, and premature mortality (e.g., Steinle et al., 2012). High spatial resolution measurements can reduce exposure misclassification and provide improved inputs for modeling. Spatially resolved air pollutant concentrations also aid in evaluating emission estimates and elucidating the effects of atmospheric processes on pollutant formation and accumulation. Urban air pollutant concentrations are known to vary by up to an order of magnitude over spatial scales ranging from meters to hundreds of meters (Marshall et al., 2008; Olson et al., 2009; Boogaard et al., 2011). Previous efforts to characterize spatial variations in air pollutant concentrations have included near-roadway sampling (e.g., Baldauf et al., 2008; Karner et al., 2010), grid-based modeling (e.g., Marshall et al., 2008; Holmes et al., 2014; Friberg et al., 2016), land-use regression models (e.g., Gilbert et al., 2005; Henderson et al., 2007; Moore et al., 2007; Marshall et al., 2008; Hankey and Marshall, 2015), satellite data (e.g., Laughner et al., 2018), dense arrays of monitors (e.g., Blanchard et al., 1999; Kanaroglou et al., 2005; Kim et al., 2018; Shusterman et al., 2018), and measurements made using mobile platforms (e.g., Brantley et al., 2014; Ranasinghe et al., 2016; Apte et al., 2017; Messier et al., 2018). The feasibility of deploying dense monitoring networks has increased with the availability of inexpensive sensors, although questions about sensor accuracy continue to be studied (e.g., Borrego et al., 2016; Castell et al., 2017; Li and Biswas, 2017; Schneider et al., 2017; Lim et al, 2019). Approaches that combine mobile monitoring with measurements made at stationary monitoring locations (Adams et al., 2012; Simon et al., 2018) or with modeling (Messier et al., 2018) are being actively researched.

The Aclima, Inc. mobile measurement and data acquisition platform was previously used with two Google Street View cars ~~and~~ equipped with research-grade instruments to measure air quality on city streets in Oakland, California between May 28, 2015 and May 14, 2016 (Apte et al., 2017) and through May 19, 2017 (Messier et al., 2018). The Oakland sampling campaign provided nearly complete coverage of all city streets with ~20 – 50 days sampling of each 30-meter (m) road segment, from which high spatial resolution maps of average air pollution concentrations were constructed (Apte et al., 2017; Messier et al., 2018). The maps reveal persistent pollution patterns with small-scale variability attributable to local emission sources; 10 – 20 driving days reproduced spatial patterns with low bias and good precision (Apte et al., 2017). The Oakland results also demonstrate the efficiency of data-based mapping: using the data from all road segments obtained on only 4 – 8 drive days represented the full data set better than did measurements from a subset of road segments combined with a land use regression – kriging model (Messier et al., 2018).

The Oakland study demonstrates an approach to mapping average air pollution concentrations within a defined geographical area by repeated sampling of each street. Mobile platform data from other locations are needed to better understand how wider coverage with more limited numbers of repeated samples within each neighborhood could be used in conjunction with data from stationary air quality monitoring locations to characterize neighborhood-scale variations. ~~For example~~In addition, new driving strategies and analytical methods could help establish concentration decay rates of mobile emissions with distance from roadways, comparability of pollutant concentrations among neighborhoods, and comparability of neighborhood concentrations to data from stationary regulatory monitors.

The mobile sampling discussed here and in Apte et al. (2017) is limited to weekdays between ~9 a.m. and 5 p.m. Sampling is necessarily conducted along roads and streets. Depending on the number of repeated driving segments, vehicles sample different road segments on different days or at different times of day. These limitations are important considerations for studies whose goal is to develop pollutant maps that represent long-term concentration averages, and which are intended to correctly characterize spatial variations at specified spatial scales. However, our study objectives are different, namely to (1) examine the capabilities of research instruments when placed in stationary and moving vehicles, (2) compare our measurements with those obtained from stationary air quality monitors, (3) evaluate driving and sampling strategies, and (4) develop statistical methods that account for sampling limitations. Limitations that are specific to our study are that (1) it was conducted as a series of geographically separated sampling campaigns between May 2016 and September 2017, generally lacking the number of repeated driving routes previously used to generate pollution maps (Apte et al., 2017; Messier et al., 2018), and (2) no set of driving routes completely covered any specific geographical domain (e.g., San Francisco or specific neighborhoods therein). The results presented here therefore focus on measurement and methodological questions that can be addressed with data available from the individual sampling campaigns. A set of research questions was developed initially and was then used to design the individual sampling campaigns. In analyzing the results, a need arose to distinguish between temporal variability (due, e.g., to sampling different places at different times) and spatial variability. Statistical methods were therefore developed to characterize spatial heterogeneity within and between neighborhoods by utilizing time-synchronized differences in the pollutant concentrations that were measured by different vehicles. Due to limited repeated sampling of individual road segments, our estimates of spatial heterogeneity do not in themselves identify locations having long-term high and low pollutant concentrations. Additional statistical methods were developed to demonstrate the use of short-term campaign measurements to characterize intermediate-scale (1 km) spatial variations of pollutant concentrations and to identify areas with short-term high pollutant concentrations, potentially indicating where more intense future sampling would be warranted.

This study examines the field capabilities of mobile research-grade instruments used in varied settings. Future work will examine the capabilities of low-cost sensor data and will address the comparability of sensor and research-grade sampler data as well as the comparability of sensors in mobile versus stationary platforms. In this manuscript, instrument measurement accuracy and precision are evaluated using weekly performance checks, laboratory audits, and independent field audits conducted through sampler inlets. The quality of the mobile instrument measurements in the ambient environment is then examined by comparisons with adjacent (<~~4~~ 9 m) stationary air quality monitoring sites and by side-by-side paired vehicle comparisons. Mobile-platform measurements are compared to data from stationary air quality monitoring sites to evaluate and validate mobile-platform data and to ensure that the mobile platforms maintain high data quality. The measurements obtained from replicate mobile platforms are compared using collocated vehicles that were operated while stationary and while driving; these results are used to establish the capabilities of the instruments for establishing high time-resolution spatial variations in pollutant concentrations. Finally, the mobile data are analyzed to examine the spatial representativeness of measurements made at stationary monitoring locations during selected time periods at a range of spatial scales (<1 km to >10 km).

The mobile measurements were made in various locations; an overview is available at https://blog.aclima.io/healthier-cities-through-data-ca-intro-6e9e22e00075 (last access, December 13, 2019). Because the driving routes were not designed to provide long-term repeated measurements for any of the locations, we did not focus on presenting pollutant maps. ~~Rather, we examined measurement capabilities and developed statistical methods for analyzing the data.~~ Data analysis methods were developed and applied to data subsets to exemplify approaches that are potentially applicable to larger data sets. Thus, some results are illustrative rather than comprehensive. Since the measurements made during the study period were intended to address specific questions based on the results from specific sampling days, analyses are presented using different subsets of the data to address different questions. While performance evaluations and audit results are documented in this manuscript for all measured species, comparisons with stationary-monitor data, between-vehicle comparisons, and summaries of spatial variations are presented only for species that were measured using more than one platform (i.e., two vehicles or one vehicle plus one stationary monitor).

2 Methods

2.1 Measurements

Measurements were made and processed by Aclima, Inc. All data are quality-assured by Aclima, Inc. at data quality levels 1 or 2 (qualified data level 1 [QD1] and qualified data level 2 [QD2]), as described in metadata documentation (Lunden and LaFranchi, 2017). The principal differences between QD1 and QD2 data are that the QD1 data include measurements made when the cars were parked overnight ~~in garages~~ and the QD2 data exclude calibration checks. Access to QD2 data is provided by Aclima, Inc. and Google, Inc. through the Google Cloud Platform using Google Cloud Shell and Google Big Query (Google, 2018). Aclima QD1 data were used for all analyses, because QD2 data (Google 2018) do not include the measurements made when the cars were parked ~~in the~~ overnight ~~parking garages~~; side-by-side comparisons of the measurements obtained when the cars were parked next to each other therefore required QD1 data sets (Aclima, 2018).

Street-level sampling was conducted in three California locations: San Francisco, Los Angeles, and smaller cities and nonurban areas within the northern San Joaquin Valley (Table 1). Measurements were made between ~ 9 a.m. and ~ 5 p.m. on weekdays, with additional sampling occurring while the vehicles were parked in the San Francisco garage and a small (~30 car) Los Angeles parking ~~garages~~ lot before (~ 6 – 9 a.m.) and after (~5 – 10 p.m.) the driving periods. The instruments were switched from vehicle to line power when parked overnight. The vehicles were parked in dedicated areas away from traffic within each overnight parking location. Specific time periods were selected for analysis to represent data from different areas and to address individual research questions (Table 2). The selected periods do not represent the full set of driving routes in any of the areas but are instead intended to ~~analyze routes that~~ address the research objectives in Table 2, as discussed ~~under results~~in Section 3. Driving routes were mapped for visualization (supplement). For clarity, data are labelled by car names (Coltrane, Flora, Rhodes; these names do not duplicate the names of any stationary monitors).

During the Los Angeles sampling, the South Coast Air Quality Monitoring District (SCAQMD) conducted through-the-inlet audits and calibration checks when the sampling vehicles were parked adjacent to stationary air quality monitoring sites (Table 3). The SCAQMD also prepared 1-minute resolution data files for measurements made at ~~various~~ these and other stationary air quality monitoring sites (Table 4; see also location map, Figure S1). Data from one of the~~se~~ dates and locations (LAXH, September 20, 2016) were suitable for collocated comparison with mobile measurements (Table 3). The stationary-monitor data from W710 consisted only of 1-hour resolution $PM_{2.5}$ mass (Table 4), which was not measured by the mobile platforms, and no data were provided for the Santa Clarita site (Table~~s~~ 3 ~~and 4~~).

The Aclima mobile measurement and data integration platform consists of fast-response (<1 s to 8 s), research-grade analyzers providing data at 1-s (1-Hz) resolution. Details about the measurement techniques along with manufacturer specifications are provided in Table S1 (see also Lunden and LaFranchi, ~~(~~2017). The inlet and sampling manifolds were designed to minimize self-sampling as well as particle and gas phase sample losses. Separate inlet lines were used for particles (copper) and gases (Teflon™, a brand name of polytetrafluoroethylene). The gas-phase inlet line was set to a 90° angle to the direction of traffic and the particle and black carbon (BC) sampling inlet line faced forward. BC was measured using a photoacoustic extinctiometer, nitric oxide (NO) was measured using chemiluminescence, nitrogen dioxide (NO_2) was measured using cavity-attenuation phase-shift spectroscopy, ozone (O_3) was measured using ultraviolet (UV) absorption, and methane (CH_4) was measured using off-axis integrated cavity output spectrometry. Particle number (PN) concentration was measured using an optical particle counter with particle counts per liter (c L^{-1}) reported in 5 size ranges: 0.3 to 0.5 micrometer (µm) ($PN_{0.3\text{-}0.5}$), 0.5 to 0.7 µm ($PN_{0.5\text{-}0.7}$), 0.7 to 1.0 µm ($PN_{0.7\text{-}1.0}$), 1.0 to 1.5 µm ($PN_{1.0\text{-}1.5}$), and 1.5 to 2.5 µm ($PN_{1.5\text{-}2.5}$).

To ensure that the 1 Hz measurements did not drift in time, on-board computers were synchronized throughout the day using Network Time Protocol (NTP), which synchronizes computers to Coordinated Universal Time (UTC) with accuracies on the order of milliseconds. Each car recorded time using ~~Network Time Protocol~~NTP and times were reported to the nearest second ~~universal time (~~UTC~~)~~. Timestamps were adjusted to account for residence time in the tubing and instrument response as described in Apte et al. (2017). We used time series plots to check the temporal comparability of vehicle and stationary monitor measurements at one-minute resolution (Section 3.3).

The gas-phase instruments received zero air and span gas weekly except for CH_4, which was checked weekly at a single concentration (2020 ppbv). Performance for the gas-phase measurements is expressed as bias and precision, defined according to the Data Quality Assessment guidelines used by the United States Environmental Protection Agency (EPA) (Camalier et al., 2007). For O_3, NO, and NO_2, the guideline analysis yields relative (in %) and absolute (in ppbv) contributions to uncertainties (Table 5). For CH_4, the analysis yields an absolute uncertainty for bias and precision of 66.7 ppbv (3.3 %), based on reference measurements at 2020 ppb.

Additional uncertainties, which range from 1 % to 3.6 %, are associated with the accuracy of the calibration gas standards and the gas delivery/generation system. Field sampling uncertainties are discussed later.

The performance of the BC and PN instruments was evaluated from collocated parked vehicles (approximately weekly for PN and nightly for BC) since certified reference standards are not available for BC and PN. Both PN and BC instruments were periodically returned to their respective manufacturers, typically once per year or when the results of ambient collocations indicated substantial drift of one car relative to the other(s) or other diagnostic checks indicated that service was required. Table 6 shows the results of evaluations performed between May 2016 and August 2017.

We calculate BC limit of detection (LOD, see footnote 2, Table 5) using data reported while the instrument is performing an internal zero, which occurs every 10 minutes for 60 seconds. This value is typically in the range of 0.2-0.3 µg m^{-3} for the 1-Hz data while the cars are parked. For vehicles in motion, we estimate 1-Hz LOD values of 0.4 µg m^{-3} for vehicle speeds less than 5 m s^{-1} and 0.8 µg m^{-3} for vehicle speeds greater than 5 m s^{-1}.

2.2 Location Uncertainty

Location uncertainty was determined as the variability of recorded positions when vehicles were parked overnight. The vehicles did not ~~have designated spaces to which they always~~necessarily return~~ed~~ to the same spaces within the disgnated Aclima parking area each night. Therefore, variances and standard deviations of parked-vehicle east-west and north-south GPS locations were determined by vehicle, date, and time of day (i.e., before and after each daily drive). Composite east-west and north-south standard deviations were then determined from individual variances weighted by sample numbers. Composite variances were converted to location uncertainty (twice the square root of the sum of the east-west and north-south composite variances). The observed 2σ location uncertainty for vehicles parked in the San Francisco parking structure was ± 6.0 m, comparable to the GPS manufacturer specifications (5 m). The location uncertainties for vehicles parked in the Los Angeles parking ~~structure~~ lot were larger (± 12.2 m at 1 s resolution and ± 11.5 m for 1-minute averages). The GPS location uncertainties therefore impose inherent limits to the spatial resolution of the data on the order of 10 m.

2.3 Comparisons between Measurement Platforms

For ambient comparisons between vehicles or between vehicles and stationary monitors, our approach for computing comparability necessarily differs from EPA guidelines for determining precision and bias, which require testing against analytical standards. B~~b~~ecause neither vehicle nor stationary monitor measurements are ~~certified target concentrations. Thus~~analytical standards, comparability must be determined in terms of the differences between measurements made by different vehicles or between vehicle and stationary-site data, which yields instrument-to-instrument comparability. Data files were merged by ~~time of day using either~~ 1-s ~~and~~ or 1-minute resolution ~~measurements. The merged data~~times and were then used to determine time-matched paired differences, which were ~~and to~~ evaluated ~~intervehicle measurement variabilities~~ as functions of ambient concentration, intervehicle distance, and vehicle speed. Paired differences were evaluated for bias of one measurement relative to another. The variabilities of the paired differences relative to the means of the paired differences were also calculated. The computational approach was necessarily limited to parameters that were measured on each of two platforms (e.g., two cars or one car plus one stationary monitor). BC and CH_4 were each measured by only one vehicle while operating (during drives, one vehicle was equipped with a BC sampler and the other with a CH_4 instrument). Therefore, it was not possible to compare BC or CH_4 concentrations between operational vehicles. A~~(a~~s previously noted, however, BC and CH_4 instruments were each installed on multiple vehicles and used to establish parked-vehicle instrument-to-instrument bias and precision: two vehicles were used in this study and two used by Apte et al. (2017)) but all four vehicles were parked in the same San Francisco garage. BC and CH_4 data were not available from stationary monitors.

2.4 Statistical Metrics

Various statistical metrics were computed to evaluate the comparability of time-paired measurements between vehicles or between vehicles and stationary monitors. These metrics include mean differences and fractional (relative) mean differences:

Mean Difference (MD) = $\mu_{A\text{-}B}$ = mean$(X_A - X_B)_i$ = Mean Difference (Car A – Car B ~~Difference~~) (1)

where $\sigma_{A\text{-}B}$ = standard error (SE) of the mean of $(X_A - X_B)_i$,

"i" denotes the "i^{th}" measurement of n paired measurements,

$SE = (\sqrt{n})^{-1} \times$ standard deviation of $(X_A - X_B)$

Fractional (relative) Mean Difference (FMD) = $\mu_{A\text{-}B} / \mu_{AB}$ (2)

= Mean Difference (Car A – Car B ~~Difference~~)/Mean of Car A and Car B Mean Concentrations

$$\sigma^2_{FMD} = \{(\sigma_{A\text{-}B} / \mu_{AB})^2 + (\sigma_{AB} \times \mu_{A\text{-}B} / \mu_{AB}^2)^2\} = FMD^2 \times \{(\sigma_{A\text{-}B} / \mu_{A\text{-}B})^2 + (\sigma_{AB} / \mu_{AB})^2\}$$

where $\mu_{A\text{-}B}$ = mean$(X_A - X_B)_i$ and $\sigma_{A\text{-}B}$ =standard error (SE) of the mean $(X_A - X_B)_i$,

$\mu_{AB} = \{(1/2) \times (\mu_A + \mu_B)\}$ and $\sigma^2_{AB} = \{(1/4) \times (\sigma^2_A + \sigma^2_B)\}$

Fractional Absolute Mean Difference (FAMD) = $|\mu_{A\text{-}B}| / \mu_{AB}$ (3)

= |Mean Difference (Car A – Car B ~~Difference~~)|/Mean of Car A and Car B Mean Concentrations

$$\sigma^2_{FAMD} = FAMD^2 \times \{(\sigma_{A\text{-}B} / \mu_{A\text{-}B})^2 + (\sigma_{AB} / \mu_{AB})^2\}$$

where $\mu_{A\text{-}B}$ = mean$(X_A - X_B)_i$ and $\sigma_{A\text{-}B}$ = SE$(X_A - X_B)_i$ and

$\mu_{AB} = \{(1/2) \times (\mu_A + \mu_B)\}$ and $\sigma^2_{AB} = \{(1/4) \times (\sigma^2_A + \sigma^2_B)\}$

Fractional Mean Absolute Difference = FMAD = $\mu_{|A\text{-}B|} / \mu_{AB}$ (4)

= Mean Difference |Car A – Car B ~~Difference~~|/Mean of Car A and Car B Mean Concentrations

$$\sigma^2_{FMAD} = FMAD^2 \times \{(\sigma_{|A-B|} / \mu_{|A-B|})^2 + (\sigma_{AB} / \mu_{AB})^2\}$$

where $\mu_{A-B} = \text{mean}|X_A - X_B|_i$ and $\sigma_{A-B} = SE|X_A - X_B|_i$ and

$$\mu_{AB} = \{(1/2) \times (\mu_A + \mu_B)\} \text{ and } \sigma^2_{AB} = \{(1/4) \times (\sigma^2_A + \sigma^2_B)\}$$

The variances σ^2_{FMD}, σ^2_{FAMD}, and σ^2_{FMAD} are derived from standard ~~statistical~~ formulae for propagating errors (~~e.g., variance of (X × Y) = {($\sigma_X \times \sigma_Y$)^2 + ($\sigma_X \times \mu_Y$)^2 + ($\sigma_Y \times \mu_X$)^2}, http://www.odelama.com/data-analysis/Commonly-Used-Math-Formulas/, last access September 27, 2019~~; Caldwell and Vahidsafa, 2019; Goodman, 1960; Ku, 1966). ~~by transforming variables (X/Z = X × Y, Y = Z^{-1}) and by making two assumptions: (1) the numerator and denominator (e.g., μ_{A-B} and μ_{AB}) are independent (implying zero covariance between differences and means), and (2) higher-order terms ($\sigma^2_X \times \sigma^2_Y$) are small compared with ($\sigma^2_X \times \mu^2_Y$) and ($\sigma^2_Y \times \mu^2_X$) (because the standard errors [σ_{A-B} and σ_{AB}] are based on large sample sizes, e.g., n > 1000, and standard errors are inversely proportional to the square root of sample size).~~ Standard errors are the appropriate measure of the variability of mean concentrations and differences, such as those defined here, whereas standard deviations are appropriately used to quantify the variability of individual measurements (see Section 3, "Results and Discussion").

The preceding equations, while expressed as car-to-car comparisons, are readily applied to other comparisons, e.g., vehicle-to-stationary monitor. If one measurement (e.g., measurement A) is defined as a reference standard, then the term μ_{AB} in the denominator of the expressions for FMD, FAMD, and FMAD may be appropriately replaced by the reference mean (μ_A). Mean differences ~~(MD)~~ are used when absolute comparisons (i.e., retaining concentration units) are informative. Fractional differences are useful for establishing vehicle-to-vehicle or vehicle-to-monitor differences relative to the magnitudes of the mean concentrations.

The FMD retains sign, i.e., indicates if $\mu_A > \mu_B$. This metric is useful when the sign is important for identifying which instrument (e.g., mobile or stationary) or which location records higher concentrations. The FAMD and FMAD are useful if the sign of the difference is not meaningful. The sign is usually not relevant, for example, in the analysis of intervehicle measurement differences as a function of the distance between the vehicles (~~see~~ "R~~r~~esults and D~~d~~iscussion"), in which the objective is to characterize the rate at which measurement comparability decays with distance. The FAMD is simply the absolute value of the FMD and both metrics approach zero when individual paired measurement differences tend to average out over a set of samples. In contrast, the FMAD provides a measure of the variability of individual measurements because it averages absolute values of concentrations. The FMAD is relevant to understanding the comparability of high-resolution (e.g., 1 s) measurements, whereas the FAMD is a measure of the comparability of a time- or space-average determined from individual measurements.

Performance audits (Tables 5 and 6) indicate that fractional differences (FAMD) exceeding ~0.1 (10 %) for gases and ~0.2 (20 %) for PN are, in general, likely to be physically meaningful relative to measurement uncertainties (bias and precision are each < 5 % for gases at concentrations > 2 – 24 ppbv; 7 – 26 % for PN and BC). Only the two largest PN size ranges exhibit bias exceeding 20 % (Table 6). Combining bias and precision indicates a total uncertainty of ~10 % for gases and ~20 % for $PN_{0.3\text{-}0.5}$. In operation, the comparability of measurements made in moving vehicles differs from those made in parked collocated vehicles (see results and discussion), so we utilize a higher threshold (i.e., 20 %) for establishing true spatial variations even for gas-phase species.

3 Results and Discussion

Mean concentrations during example study periods are summarized in Table 7 for context. Subsequent analyses of spatial heterogeneity, which are presented in later subsections ~~in this section, which~~ and depend on the availability of measurements from two or more sampling platforms, focus on NO, NO_2, O_3, and $PN_{0.3-0.5}$. These pollutants are of interest because they are measured with differing accuracies, they exhibit differing degrees of spatial variation, and they vary in their degree of atmospheric chemical processing. NO is a primary pollutant and NO_2 forms rapidly (i.e., minutes) from NO. NO_2 formation and O_3 loss are linked through the rapid reaction of NO with O_3 to form NO_2; Seinfeld and Pandis (2016) calculate a 1/e lifetime for NO of 42 seconds at 50 ppb O_3. O_3 formation and accumulation occurs more slowly (i.e., hours) from NO_2 and volatile organic compounds (VOCs) in the presence of ultraviolet (UV) radiation (Seinfeld and Pandis, 2016). $PN_{0.3-0.5}$ is the smallest size fraction that was measured, present in the highest numbers (83 % of PN, Table 7), and is likely indicative of newly aged particles from fresh motor-vehicle emissions (Zhang and Wexler, 2004; Zhang et al., 2004; Zhu et al., 2002).

The fraction of PN in the 0.3 – 0.5 μm size fraction was lower in spring (60 % in San Francisco, May 2017 and 72 % in the San Joaquin Valley, March 2017) and higher in summer (90 % in Los Angeles, August 2016) and autumn (86 % in Los Angeles, September 2016 and 84 % in the San Joaquin Valley, November 2016) (Table 7). Although these differences in the PN size distributions possibly reflect regional-scale spatial variability, no simple comparison among regions is possible due to sampling them during different seasons. ~~they more likely~~The regional differences could in fact reflect seasonal variations in PM composition: the observed variations in PN distributions are consistent with past studies that indicate the importance of PM nitrate (NO_3) found in larger (> 0.5 μm) size fractions primarily as ammonium nitrate in California during cooler months (e.g., Herner et al., 2005), which could lead to the observance of different size distributions in the different regions.

Mean concentrations of gases were comparable among the study locations and periods (Table 7). O_3 concentrations were highest in Los Angeles in August near downtown (south of the CELA site, Figures S6 and S7) followed by concentrations in September in west Los Angeles near the WSLA site (Figure S8) and near Los Angeles airport (near the LAXH site, Figure S3). Mean O_3 in the remaining locations (SJV and SF) fall within a narrow range (23 – 29 ppbv) and are only a factor of less than 2 lower than in Los Angeles. Mean concentrations of NO_2 also vary by a factor of 2 with highest concentrations near the LA airport and lowest concentrations in SF (Table 7). Concentrations of NO are highest by a factor of about 2 in Los Angeles near the airport and in the SJV in November during mostly freeway driving. At all locations studied, typical NO-NO_2-O_3 chemistry was observed with higher NO and NO_2 concentrations and lower O_3 levels near mobile emission sources. Mean methane concentrations were low (~ 2 ppmv) during all periods and varied among areas within <0.1 ppmv. As with PN, these average concentrations likely vary due to time of year, location relative to source emissions, and chemical processing.

3.1 ~~How Well Do~~Comparability of Measurements in the Mobile Platforms ~~Compare~~ to the Inlet Audits~~?~~

Field c~~C~~alibration checks (zero and span) were conducted ~~in the field~~through inlets using SCAQMD equipment and standards; these checks were compared with Aclima calibration checks that were made before, during, and after the period when vehicles drove in ~~the~~ Los Angeles (Table 8). The SCAQMD and Aclima checks were comparable and indicate that measurements of the tested gas-phase species (NO, NO_2, and O_3) maintained accuracy and replicability in the field during the Los Angeles driving routes. The Los Angeles drives followed the same field protocols as the drives in San Francisco and the San Joaquin Valley. The cross-lab differences between the Aclima and SCAQMD calibration checks (defined as the lab-to-lab differences in the mean relative differences from target concentrations averaged over all calibration checks) were -5 % ± 2.0 % for NO, -1.5 % ± 1.0 % for NO_2, and +0.5 % ± 1.3 % for O_3 (not tabled). All differences were less than the invalidating limits for the South Coast Air Quality Management District's weekly calibration checks: 7% for O_3 and 10% for CO, SO_2, and NO_x (Table 2.4, https://ww3.arb.ca.gov/aaqm/qa/pqao/repository/district_sops/south_coast/quality_assurance/qapp_criteria_pollutants.pdf, last access April 15, 2020).

3.2 ~~How~~ Similarity ~~are~~ of Concentrations Obtained from Collocated Vehicles when Parked and w~~W~~hen Moving~~?~~

Car-to-car comparisons were made to evaluate the comparability of collocated ambient measurements made while the vehicles were parked and while driving (Table 9). The cars generally followed different routes, as discussed later; when the cars travelled a route segment together, they drove "caravan style", keeping each other in sight but not following immediately one behind the other. ~~These~~ Time-synchronous measurement differences reflect a combination of instrument and ambient sampling uncertainties; for moving vehicles, differences may also reflect spatial variability, depending on measurement integration times relative to intervehicle distances. The comparisons are expressed as mean car-to-car differences plus-or-minus 1 standard deviation of the paired 1-s differences, yielding metrics for car-to-car measurement bias and variability, respectively, averaged over ~1000 – 50,000 paired differences.

The observed mean paired differences between parked vehicle measurements were 0.2 – 3.9 ppbv for NO, 0.3 – 1.9 ppbv for NO_2, and 0.8 – 4.5 ppbv for O_3 (Table 9). The corresponding FAMD (absolute values of mean differences divided by mean concentrations) range from 0.03 – 0.24 (3 % to 24 %) for gases and 0.04 – 0.22 (4 % to 22 %) for PN. These differences are comparable to, or larger than, instrumental bias and precision (<5 % each for gases at concentrations > 2 – 6 ppbv, Table 5; 10 – 11 % for $PN_{0.3-0.5}$, Table 6). For gases and PN, the variabilities (standard deviations) of the 1-s paired differences exceed the mean differences (except O_3 during the SJV sampling period of November 16 – 23, 2016), which is expected because instrumental variations average toward zero when instruments are unbiased with respect to each other. The mean paired differences varied among individual sampling days (Figure S2). Between-vehicle 1-s variability is higher in closely-spaced moving vehicles than in stationary vehicles, especially for NO_2 (Table 9; note that this comparison could not be made for NO).

We interpret this difference as indicating that moving vehicles sampled heterogenous parcels of air and the intervehicle measurement differences are thus due to fine-scale spatial variability.

3.3 ~~How~~ Similarity of ~~are~~ Mobile ~~Reference~~ Concentrations to Stationary Monitor Data~~?~~

For field comparisons to stationary monitors, we worked with SCAQMD staff who operate the monitors and are familiar with all measurements made at each location. On September 20, 2016, two sampling cars parked next to the monitor at LAXH (Tables 3 and 4, Figures S3 and S4). Relative to the ground-level position of the stationary monitor probe (located inside a fenced enclosure), the vehicles alternated positions from closer when audited (Coltrane 6.6 m from LAXH, Flora 8.5 m from LAXH) to further when sampling (Coltrane, 24.1 m for 1 hour; Flora, 18.5 m for 2 hours) as determined from GPS coordinates for the monitor and vehicles. The heights of the LAXH instrument probes are 4.2 m above ground level (SCAQMD, 2018a), whereas the mobile sampler inlet heights are 2 m above ground level. The monitoring instruments at LAXH are in a vacant field north of Los Angeles International Airport (Figure S4). The site is surrounded by several schools to the NE, N, and NW with residential communities (Playa Del Rey and Westchester) north of the airport and further away surrounding the site. The closest communities include homes and 2 – 4 story apartments. Minimal traffic is expected immediately adjacent to the site. The mobile platforms recorded mean concentrations of NO, NO_2, O_3, and O_x (= NO_2 + O_3) that were comparable to LAXH monitor concentrations: most mean paired differences between mobile-platform and LAXH concentrations were less than 10 % of the average concentrations (Table 10). Time series of 1-minute Flora, Coltrane, and LAXH measurements show agreement (Figure S5) (mean Flora – Coltrane distances were 12.2 and 20.2 m). CH_4 ~~concentrations can be a potential tracer of~~ is reported in ~~fresh~~ motor-vehicle emissions (Nam et a., 2004), so a correlation between NO and CH_4 will usually be observed when sampling fresh automotive exhaust emissions; ~~and~~ all NO values correlated with Coltrane CH_4 concentrations (r^2 = 0.84 to 0.87; Flora did not report CH_4).

3.4 ~~How Large are the~~ Differences between Mobile ~~Reference~~ Concentrations and Stationary Monitor Data w~~W~~hen the Cars are Not Close to Monitors~~?~~

Spatial variation is defined by differences in time-synchronous measurements made in differing areas. To interpret the paired differences as spatial variation, rather than measurement uncertainty, we refer to the preceding analyses of instrument and sampling performance in audit tests (Tables 5 and 6) and collocated vehicles (Table 9). As previously noted, the results for measurement bias and precision (Tables 5 and 6) and for comparability of collocated vehicles (Table 9) lead us to define FAMD > 0.2 (20 %) as an indicator that spatial variations exceed measurement and sampling uncertainties. The intent of the analyses in this section is to help elucidate the spatial scales over which stationary-monitor and mobile-platform data represent ambient concentrations and to characterize spatial heterogeneity of pollutant concentrations within neighborhoods.

Because vehicles sampled different road segments on different days and at different times of day, we compiled time-synchronous differences between the concentrations measured by two cars (or cars and monitor) to remove the confounding effects of day-to-day and diurnal variability. Random differences, such as short, intermittent exposures of one car to a highemitting vehicle or to variations in wind directions, are averaged out in the FAMD statistic. In contrast, systematic car-to-car (or car-to-monitor) differences yield higher FAMD values. Systematic differences could occur if the instrumentation in one car was biased relative to the other car (e.g., Apte et al., 2017) or to the monitor. If instrumental sources of systemic car-to-car or car-to-monitor difference can be eliminated through side-by-side sampling comparisons (Sections 3.2 and 3.3), we can then conclude that larger FAMD values (e.g., > 0.20 or 20%) represent spatial heterogeneity due to the two cars sampling different neighborhoods. FAMD is also a useful metric for evaluating the spatial scale of representativeness of stationary monitors. The relationships between FAMD and vehicle-monitor or intervehicle distance, discussed below, characterize the spatial scales of pollutant heterogeneity but do not indicate which neighborhoods experienced higher pollutant concentrations. For that purpose, we examined maps (Section 3.4) and developed the visualization discussed in Section 3.5.

3.4.1 Los Angeles, August 2016

Between August 3 (the first complete Los Angeles driving day) and August 12, the two vehicles traversed different neighborhoods south of the central Los Angeles stationary monitor (CELA, Table 4; Figures S6 and S7) ~~at varying speeds~~ between 9 a.m. and 6 p.m. at car-monitor distances ranging from 1 to 7 km~~ (Figure 1)~~. The monitoring instruments at CELA are located on a rooftop of a two-story building and the heights of various instrument probes range from 11 to 12 m above ground level (SCAQMD, 2018b). Driving routes for the first sampling day (August 3) are shown in Figure S6; most of the routes on other dates were similar. In general terms, the US-101 and one section of the I-5 freeways run across the southern border of the sampling area; the area sampled is split by a N-S portion of I-5 and bordered on the north by I-10. The I-10 freeway is situated between CELA and the measurement area. For comparison with the 1-minute resolution CELA data, 1-minute average concentrations were created from the 1-s mobile-platform data. Because driving speeds averaged ~ 2 – 5 m s^{-1}~~ (Figure 1)~~, the typical distances travelled in one minute were ~100 – 300 m. The 1-minute average positions of the mobile sampling are visibly discrete (Figure S6). Differences between CELA and car 1-minute concentrations were highest when cars drove along freeways but also show spatial heterogeneity within the neighborhoods sampled (Figure 1). While in motion, generally beginning after 9:00 a.m. and ending between 5:00 and 6:00 p.m., the cars recorded higher concentrations of NO and NO_2 than the CELA stationary air monitor did, likely due to the proximity of fresh vehicle emissions experienced by street-level sampling in the vehicles (Figures 1 and 2). During the driving hours, the vehicles recorded lower levels of O_3 than CELA did (Figure 2~~1~~). As noted in the previous comparison of collocated and stationary-monitor data, much of this difference is attributable to street-level reaction of fresh NO emissions with O_3; this interpretation is supported by the closer agreement between cars and CELA of O_x than O_3 (Figure 2~~1~~).

To quantify differences within and between neighborhoods, b~~B~~etween-vehicle paired comparisons were determined as differences between time-synchronous 1-min mobile concentrations for August 3 – 12 (near CELA), which were then averaged over 0.5 km bins (0 – 0.25 km, 0.25 – 0.75 km, etc.) (Figure 3~~2~~). The bin-average FAMDs ranged from 0.02 (2 %) at 0.125 km to 0.14 – 0.44 (14 – 44 %) at 4.5 – 5.5 km (mean = 0.12, or 12 %, over all bins) for NO_2 and from 0.006 (0.6 %) at 0.125 km to 0 – 0.07 (0 – 7 %) at 4.5 – 5.5 km (mean = 0.02, or 2 %, over all bins) for O_3. For these two pollutants, the mean differences among streets and neighborhoods were therefore small (12 % and 2 %, respectively, at 0.125 – 5.5 km spatial scale). For NO, bin-average FAMDs were larger and ranged from 25 % at 0.125 km to 4 – 75 % at 4.5 – 5.5 km.

The intervehicle differences averaged over distance bins concisely summarize large numbers of measurements but this averaging could mask finer spatial variations of possible interest. ~~The results obtained for bin averages were examined for higher variability on smaller spatial scales.~~ We compared the standard deviations of the mean intervehicle concentration differences to the corresponding mean concentrations to characterize variability within the spatial averages. These ratios (standard deviation of intervehicle difference/mean concentration) ranged from 0.4 to 1.0 (average = 0.5) for NO_2. Within the binned intervehicle averages, therefore, vehicle-to-vehicle NO_2 concentration differences varied by up to a factor of two (twice the standard deviation of the mean differences) times the mean observed concentrations. For NO, the ratios ranged from 1.2 to 4.0 (average = 2.8), indicating that vehicle-to-vehicle NO concentration differences varied by up to a factor of six (two standard deviations) within the binned intervehicle averages.

The number of particles in the size range 0.3 to 0.5 μm exhibited FAMDs exceeding 0.2 (20 %) that were less variable than the NO FAMD. Both NO concentrations and particle numbers likely varied as the vehicles sampled different streets and neighborhoods and experienced differing levels of fresh emissions at any given time (e.g., Figures S6 and S7). The peak in the NO FAMD at 3 and 3.5 km corresponds to mean NO concentrations of 6.6 and 8.1 ppbv, respectively for Flora and mean NO concentrations of 14.8 and 15.3 ppbv, respectively, for Coltrane. Many of the 85 and 120 1-minute differences in these two bin averages correspond to cases where Coltrane sampled close to the confluence of the Santa Anna and Golden State freeways while Flora collected data further from freeways (Figure S7). An approach to identifying high-concentration locations is illustrated later in the discussion of data from San Francisco (Section 3.5).

The NO FAMD for car-CELA comparisons largely exceeded 1; the NO_2 and O_3 FAMDs were less than 0.5 and 0.2, respectively, at most car-CELA distances (Figure 4~~3~~). Although the two cars drove different routes, the two car-CELA comparisons were similar (Figure 4~~3~~). The representativeness of CELA and other sites is discussed below (Section 3.6).

3.4.2 Los Angeles, September 2016

Driving routes were near (<0.~~5~~2 to 5 km) the west Los Angeles stationary monitor (WSLA, Table 4) on four of the 14 days between September 12 and 30 (including areas shown in Figure S8 for September 13 and 19; similar routes were driven on September 26 and 29). Drives began at ~9 a.m. and ended by 5 p.m. LDT. Because only one car drove near WSLA on each of the four days, only car-to-WSLA comparisons are presented. The monitoring instruments at WSLA are located on the roof of a trailer on the grounds of the VA hospital and the heights of the instrument probes are 4.2 m above ground level (SCAQMD, 2018c) (Figure S9). The monitor is located <600 m west of I-405 and about 200 m south of a major arterial, Wilshire Blvd. The immediate surrounding area to the north and south is grass with some trees, and slightly further out the area is primarily residential multistory (2 – 3 stories) apartment buildings.

The mobile platforms recorded substantially (between 70 % up to a factor of 32) higher concentrations of both NO and NO_2 than WSLA while the cars drove from the parking garage on the Santa Monica freeway to the neighborhood destinations

(Figure 5~~4~~, WSLA-car distances > 5 km). Even at distances < 0.2~~5~~ km up to 5 km from WSLA, the mobile platforms recorded higher concentrations of NO and NO_2. However, mean car and WSLA O_x concentrations at distances < 10 km were more similar than were corresponding car and WSLA concentrations of NO_2 and O_3 (Figure 5~~4~~). For NO and NO_2, the FAMD exceeded 1.5 and 0.4, respectively, at all distances outside the parking garage (Figure 6~~5~~). During part of their routes, the cars sampled adjacent to the San Diego (I-405) freeway, which likely contributed to higher mean NO and NO_2 concentrations for the mobile platforms. The WSLA monitoring site (grounds of VA hospital) has a middle scale zone of representation (100 m to 0.5 km) for NO_2 (Table 4), consistent with our results. For O_3 and O_x, the FAMD were < 0.2 and < 0.05, respectively, within 5 km of WSLA.

3.5 ~~How Large are the~~ Differences between Pollutant Concentrations Reported by Vehicles Operating in Different Neighborhoods~~?~~

~~Answers to this question~~This section helps identify neighborhoods where pollutant concentrations are typically higher than occur elsewhere, potentially indicating where long-term monitors could be located for characterizing higher pollution impacts. In such neighborhoods, air pollutant exposures ~~are potentially~~may be higher than levels measured by regulatory monitors, since the latter are typically focused on community-scale air pollution. ~~Analyses are useful for identifying areas experiencing higher pollutant concentrations and, potentially, locating long-term monitors for characterizing higher pollution impacts.~~

3.5.1 San Francisco, May 2017

Measurements made by paired vehicles operating in different neighborhoods of San Francisco between May 1 and 12, 2017, are used to illustrate short-term (two week) neighborhood-scale spatial variability. Example driving routes are shown as 1-s averages for one day in Figure S10. The 1-s data were aggregated to 1-minute averages and the one-minute averages for all routes for May 1 - 12 are depicted in Figure 7~~6~~a. Different routes were taken on different days to obtain measurements in different neighborhoods in San Francisco. Since the averaging driving speeds between May 1 and 12 were 4.5 and 4.8 m s^{-1} for Coltrane and Flora, respectively, the positions shown in Figure 7~~6~~a represent the midpoints of segments averaging 270 – 290 m.

One-minute averages were next averaged spatially to the nearest kilometer (based on conversion of latitude and longitude to Universal Transverse Mercator [UTM] coordinates) separately for each car (Figure 7~~6~~b), which is a spatial scale corresponding to about a 3-minute average. However, the sampling times of the 1-km average concentrations varied by up to six hours among locations, which confounds spatial with diurnal variability. Instead of analyzing 1-km average concentrations by vehicle, therefore, each 1-minute average was paired with the corresponding 1-minute average reported by the other vehicle and synchronous concentration differences were determined. When these synchronous differences are averaged to 1-km resolution, they represent the average enhancement or deficit of a pollutant at a given 1 km location when compared to simultaneous measurements made elsewhere, i.e., the average excess or deficit relative to co-measured concentrations (Figure 7~~6~~c and 7~~6~~d).

This approach permits consideration of spatial variations in a manner that limits the confounding influence of diurnal variability and provides a better relative comparison of pollutant levels among neighborhoods.

One-km averages consisting of fewer than ten 1-minute data points were excluded, yielding 97 of 236 possible spatial averages for NO_2 and 107 of 271 possible spatial averages for O_3. The decision to exclude 1-km averages consisting of fewer than ten 1-minute data points was based on the high standard errors of such averages (e.g., > 0.2 for the NO_2 FAMD when n < 10). The number of 1-minute averages within each 1-km average ranged from 10 to 95 (i.e., 60 – 5700 1-s averages); for, the 1-km average covering the parking garage, there were 1813 and 2520 1-minute O_3 and NO_2 averages, respectively.

For both NO_2 and O_3, most 1-km average concentration differences exceeding 2 ppbv (or < - 2 ppbv) were statistically ~~significant~~ nonzero (i.e., the interval of the mean difference ± 2 standard errors of the mean did not cover zero); most differences in the range between -2 and 2 ppbv were not statistically different from zero (Figure 7~~6~~c and 7~~6~~d). These figures exclude the few larger differences that were not statistically different from zero (7 O_3 and 4 NO_2 averages), which may include atypical events. Both fractional differences and the signs (excess or deficiency) of the differences are of interest; therefore, the mean fractional differences are expressed as FMD rather than FAMD (Figures 7~~6~~e and 7~~6~~f) since the sign of the difference is important. For NO_2, FMD exceeding 0.5 (or < -0.5) were statistically different from zero; for O_3, FMD exceeding 0.05 (or < -0.05) were statistically different from zero. The contrast in the detectability of ~~statistically significant~~nonzero fractional NO_2 and O_3 differences between vehicles (FMD) is pronounced but readily explained: the average intervehicle concentration differences were comparable for NO_2 and O_3 (Figure 7~~6~~c and 7~~6~~d), but mean O_3 concentrations exceeded mean NO_2 concentrations (Table 8).

During May 1 – 12, locations on the east side of San Francisco experienced higher NO_2 concentrations and lower O_3 concentrations than central and western locations (Figure 7~~6~~). This result is consistent with typically prevailing winds from the west to northwest and with high traffic volumes on major freeways, I-80 (Bay Bridge), I-280, and US 101, which are expected to yield higher emissions and ambient concentrations closer to areas with higher traffic volumes. Because fresh NO emissions initially reduce ambient O_3 concentrations, O_3 concentrations are typically lower where NO_2 concentrations are higher. The results of this limited analysis indicate that the measurement system can reveal differences among air pollutant levels occurring in different neighborhoods during short (i.e., days to weeks) time periods.

The San Francisco results reveal mean 1-km scale ~~enhancements~~ spatial differences ~~(FAMD)~~ in NO_2 and O_3 concentrations up to 117 % and 46 %, respectively, of mean values during the two-week sampling period. The results obtained for 1-km averages can be further examined to demonstrate higher variability on smaller spatial scales. We compared the standard deviations of the 1-km mean intervehicle NO_2 differences to the corresponding 1-km mean NO_2 concentrations to characterize variability within 1-km spatial averages. These ratios (standard deviation of intervehicle difference/mean concentration) ranged from 0.5 to 3.0 (average = 1.3). Within the 1-km averages, therefore, vehicle-to-vehicle NO_2 concentration differences varied by factors of 1 – 6 (twice the standard deviation of the mean differences) times the mean observed 1-km average concentrations. Another indicator of spatial variability at finer resolution is the FMAD: as previously noted, the FMAD provides a measure of the variability of individual measurements because it averages absolute values of concentrations and is therefore relevant to understanding the comparability of high-resolution measurements. For the San Francisco data, the FMAD represents the variability of the 1-minute time averages that comprise each 1-km spatial average. The average of the FMAD values across all 1-km spatial averages was 0.74, nearly twice as high as the average FAMD of 0.44.

3.5.2 San Joaquin Valley, November 2016

Over ten months, driving routes in the northern San Joaquin Valley were located within the cities of Tracy (2017 population 90,890), Stockton (320,554), Manteca (76,247), Merced (84,464), Modesto (215,080), and Turlock (72,879) (https://www.cacities.org/Resources-Documents/About-Us/Careers/2017-City-Population-Rank.aspx, last access December 2, 2019) (Table 1). The initial drives occurred November 16 – 23, 2016 (Figure 8~~7~~; see also examples of drives on other days in Figures S11 – S15). Because the destinations were located over 100 km from where the cars were parked overnight in the San Francisco parking garage, the cars drove longer distances and sampled more non-urban roads (both rural and high-traffic volume interstates) each day than they did in Los Angeles or San Francisco. The San Joaquin Valley car-to-car comparisons therefore provide insight into variations on larger spatial scales (e.g., 10 – 100 km), which are of interest for understanding enhancements of urban over non-urban pollutant concentrations as well as pollutant transport between cities or subregions.

Between November 16 and 23, 2016, the cars drove on non-urban roads and on city streets in Stockton, Manteca, and Modesto, providing information on pollutant concentrations in Stockton relative to other portions of the northern San Joaquin Valley and in the eastern half of the San Joaquin Valley compared with the western side (Table 11; Figures 7 and S11 – S15). For each geographical pairing, pollutant enhancements varied by pollutant and date (Table 12; see Tables S1 – S4 for detailed tabulations). For example, relative to sampling in both a rural area and near I-205 in Tracy, Stockton exhibited enhancements of NO_2 concentrations and $PM_{0.3\text{-}0.5}$ counts on November 16 along with deficits of NO and O_3. Since mean NO_x (NO + NO_2) concentrations in Stockton (31.3 ppbv) did not differ from the rural route (31.8 ppbv) (Tables S1, S2), the Stockton – rural differences in NO and NO_2 concentrations may have been related to atmospheric chemical reactions and air mass aging. On November 23, the Stockton – highway comparison exhibited the opposite pattern to November 16: deficits of NO_2 concentrations and $PM_{0.3\text{-}0.5}$ c L^{-1} along with enhancements of NO and O_3 (Table 12) compared to routes in Modesto (within 1 km of Highway 99) and along Highway 99 (Modesto to Merced), Highway 140 (Highway 99 to I-5), and I-5 (Figures S15). High traffic volumes (~50,000 – 150,000 vehicles per day, annual average peak volumes) are typical of Highway 99 (https://dot.ca.gov/programs/traffic-operations/census/traffic-volumes, last access April 15, 2020), so the results on ~~this date~~ November 23 indicate higher pollutant concentrations on and near major highways than on city streets in Stockton and in Modesto (Tables 12, S2 – S5~~1~~).

The spatial analyses do not show consistent enhancements of pollutant concentrations in northern San Joaquin Valley cities over concentrations occurring in surrounding areas. This result suggests a complex situation in which pollutant levels in the study cities depend on both local emissions and intra-regional pollutant transport. Similarly, the relationships between measured concentrations and intervehicle distance in the San Joaquin Valley depend upon the locations of the vehicles (Figure S16). Results for November 16 are shown for multiple species in Figure S17. NO_2 and particle numbers exhibited FAMDs exceeding 0.2 over most intervehicle distances. The largest FAMDs for NO_2 and particle numbers were associated with contrasts between locations within the San Joaquin Valley and locations along an upwind boundary; these contrasts appear as intervehicle distances of 50 – 80 km, corresponding to times when Coltrane traversed the highway between San Jose (hour 11) and Crows Landing (near hour 14 at I-5 in the San Joaquin Valley) while Flora was sampling city streets in Stockton (Figure 7). Paired O_3 values were similar (FAMD < 0.2 up to intervehicle distances of 50 km), illustrating the regional character of O_3 in much of the northern San Joaquin Valley. The smaller FAMDs at 25 and 45 km intervehicle distances occurred when both vehicles were sampling freeway locations in the urban San Francisco Bay area (Figure S17). The larger FAMDs at intervehicle distances of 15 km occurred when the cars traversed I-580 between Manteca and Hayward (near Castro Valley Freeway, Figure 8~~7~~) on their return trip in the afternoon and the vehicles experienced differences in traffic levels due to their positions in urbanized versus nonurban portions of I-580 (hour 15, Figures 8~~7~~, S17).

3.6 ~~How~~ Spatial~~ly~~ Representation~~ve~~ ~~are~~ of Measurements from Regulatory Monitors~~?~~

Comparisons of mobile-platform concentrations to concentrations recorded by the downtown Los Angeles stationary monitor (CELA) showed that the FAMD for NO largely exceeded 1 (100 %); ~~the~~ most NO_2 and all O_3 FAMDs were less than ~~0.5 (50 %) and~~ 0.2 (20 %)~~, respectively,~~ at car-monitor distances ranging from 0.5~~1~~ to ~~7~~ 4 km. The results indicate that the U.S. EPA classification of the downtown Los Angeles location as a neighborhood scale site (0.5 – 4 km zone of representation, Table 3) is appropriate for NO_2 and O_3. Comparisons of mobile monitors to data from the west Los Angeles monitor (WSLA) showed that the mobile platforms recorded much higher concentrations of NO and NO_2 than the monitor at vehicle-to-monitor distances ranging from < 0.5 km to 5 km; for NO and NO_2, the FAMD exceeded 1.5 (150 %) and 0.6 (60 %), respectively. The results support the U.S. EPA classification of WSLA as a middle scale site (100 m to 0.5 km zone of representation, Table 3). The methods used for evaluating the spatial representativeness of CELA and WSLA are readily applied to other locations.

3.7 ~~How~~ Effectiveness of ~~Were~~ the Driving Routes for Addressing Study Questions~~?~~

~~The d~~Driving routes ~~that were~~ followed in this study were intended to address various research questions focused on evaluating mobile platform performance and spatial scales of representativeness (per previous subheadings in "Results and Discussion"). Different routes were deployed for different questions. The routes utilized in the comparisons with stationary regulatory monitors in Los Angeles provided effective coverage of neighborhoods located 100 m to 4 km from two stationary monitors. The results supported the EPA classifications of those monitors.

The sampling conducted in San Francisco was intended to delineate spatial variations of pollutant concentrations across the city. Sampling during a single two-week period, which covered a subset of a compact urban environment, clearly revealed 300 m – 1 km spatial differences in pollution concentrations but varied by pollutant. In contrast, sampling was conducted over a much larger area in the northern San Joaquin Valley and the results were difficult to interpret from a limited (two-week) set of measurements because the spatial domains sampled were different on different days. For example, contrasts between an urban area (Stockton) and areas surrounding Stockton were expected to yield information on the urban pollution enhancement in

Stockton. However, three different types of environments were sampled in conjunction with the initial two weeks of Stockton measurements: (1) nearby cities (e.g., Manteca, Tracy, and Modesto, located 19 to 45 km from Stockton), (2) a major freeway (Highway 99, mean distance 61 km from Stockton), and (3) a rural area (56 km from Stockton). Establishing quantitative contrasts for each of these comparisons likely requires at least two weeks of data for each type of comparison (e.g., Stockton vs rural). Such comparisons could be explored using the full San Joaquin Valley data set.

4 Conclusions

The Aclima, Inc. mobile measurement and data acquisition platform, which equips Google Street View cars with research-grade instruments to measure air quality at high spatial resolution, is an effective approach to obtaining improved understanding of spatial variations in air pollutant concentrations. Data provided by the system will be highly useful for evaluating air quality management policies intended to reduce human air pollutant exposure, acute and chronic health impacts, and premature mortality. Audit results demonstrate that reference instruments in stationary vehicles are capable of reliably measuring NO, NO_2, O_3, and PN with bias and precision ranging from <5 % to <25 % at 1-s time resolution.

During experiments conducted in Los Angeles, San Francisco, and the San Joaquin Valley, California, collocated parked and moving mobile platforms replicated mean NO, NO_2, O_3 concentrations with mean differences in 1-s measurements ranging from 0.2 to 5.6 ppbv; mean differences in $PN_{0.3\ \mathrm{to}\ 0.5}$ varied from 500 to 21,000 c L^{-1}. On a relative basis, the mean differences between replicate mobile platforms ranged from 1 % to 37 % of the mean NO, NO_2, and O_3 concentrations and 2 % to 32 % of PN, with higher mean differences observed in the larger particle size ranges (which also had few numbers of particles). The majority (21 of 26) comparisons of collocated mobile platforms exhibited differences <20 % of the mean concentrations, thereby suggesting that differences exceeding 20 % obtained by vehicles operating simultaneously in different neighborhoods represented measurable spatial variation.

Paired time-synchronous mobile measurements were used to characterize the spatial scales of concentration variations when vehicles were separated by <1 to 10 km. Measurements made in Los Angeles during August 2016 exhibited intervehicle FAMD that ranged from 2 % at 0.125 km to 14 – 44 % at 4.5 – 5.5 km (mean 12 %) for NO_2 and from 0.6 % at 0.125 km to 0 – 7 % at 4.5 – 5.5 km (mean 2 %) for O_3. The standard deviations of bin averages indicated that finer-scale (e.g., 100 – 300 m, 1-minute averages) intervehicle variations were larger, indicating variability by up to a factor of two for NO_2 and a factor of six for NO (two standard deviations) within the binned intervehicle averages.

For NO and $PN_{0.3-0.5}$, bin-average mean differences exceeded 20 % for the same driving routes, indicating measured spatial variability exceeding the uncertainties in measurement methods when employing the mobile platforms. For NO, the standard deviations of bin averages ranged from 1.2 to 4.0 (average = 2.8), indicating that vehicle-to-vehicle NO concentration differences varied by up to a factor of six (two standard deviations) within the binned intervehicle averages.

A data analysis approach was developed to characterize spatial variations in a manner that limits the confounding influence of diurnal variability. The approach involved examining synchronous differences between 1-minute measurements made by two mobile platforms, which were then averaged to one-kilometer resolution. The approach was illustrated using data from San Francisco, revealing mean 1-km scale ~~enhancements~~ spatial differences in NO_2 and O_3 concentrations up to 117 % and 46 %, respectively, of mean values during a two-week sampling period. Within the 1-km averages, vehicle-to-vehicle NO_2 concentration differences varied by factors of 1 – 6 times the mean observed 1-km average concentrations, implying higher variability at spatial scales <1 km (i.e., among 1-minute averages, corresponding to ~300 m distances). Locations on the east side of San Francisco experienced higher NO_2 concentrations and lower O_3 concentrations than central and western locations likely due to differences in traffic density and to meteorological factors, with prevailing winds from the west or northwest.

The mobile data were also used to provide insight into the spatial representativeness of measurements made at stationary monitoring locations. Comparisons of mobile measurements to data from two stationary monitors in Los Angeles indicate that the U.S. EPA classifications of the monitors as representative of neighborhood (0.5 – 4 km) or middle (100 m – 0.5 km) scale pollutant concentrations are appropriate. The methods used for evaluating the spatial representativeness of the two monitors are readily applied to other locations.

5 Data Availability

Access to Aclima QD2 data is provided by Google, Inc. on request (https://goo.gl/EJMcCD) through the Google Cloud Platform using Google Cloud Shell and Google Big Query ~~(https://bigquery.cloud.google.com/table/street-view-air-quality:California_201605_201709_GoogleAclimaAQ.California_2016_2017?tab=details&pli=1).~~ (https://bigquery.cloud.google.com/table/street-view-air-quality:California_201605_201709_GoogleAclimaAQ.California_2016_2017?tab=details&pli=1).

6 Contributions

All authors contributed to the manuscript. P. S. and D. V. established science questions to be addressed in consultation with Aclima, Inc. and EPA staff. P.S coordinated the project for EPA through December 2018 and continued to work on the project after leaving EPA and while currently serving as a consultant to Aclima Inc. M. L. and B. L. managed the project for Aclima, Inc., supervised driving routes, evaluated measurement accuracy, and compiled data sets. C. B. and S. S. carried out analyses of the data sets. C.B. wrote the manuscript with contributions from each co-author.

7 Competing Interests

The authors declare that they have no conflict of interest. M. L. and B. L. are employed by Aclima, Inc. P.S. serves as a consultant to Aclima, Inc.

8 Disclaimer

Sections or the underlying technologies described herein are proprietary, owned by and subject to the intellectual property rights of Aclima, Inc. All rights reserved. The views expressed in this article are those of the authors and do not represent the views or policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement, certification, or recommendation for use.

9 Acknowledgements

Funding for this project was provided by the EPA Office of Research and Development through a cooperative agreement with the Electric Power Research Institute (Coop No. 83925001-0). In-kind support was provided by Aclima and EPRI. Aclima collected the data, provided quality assurance, and provided access to the data used in this study. Mike Hamdan of the South Coast Air Quality Management District generously facilitated field comparisons with stationary monitors. We thank Surender Kaushik who became the project manager after Paul Solomon retired from EPA. We also gratefully acknowledge K. Tuxen-Bettman, D. Herzl, O. Puryear, the Aclima mobile platform team, and the Google Street View team and drivers for their contributions to the project.

10 References

Aclima EPA Science Questions. https://console.cloud.google.com/storage/browser/aclima_epa_science_questions?pli=1 (last access, October 7, 2019; access permission granted by Aclima), 2018.

Adams, M.D., P.F. DeLuca, D. Corr, and P.S. Kanaroglou: Mobile air monitoring: measuring change in air quality in the city of Hamilton, 2005-2010, Soc. Indic. Res. 108, 351-364, doi: 10.1007/s11205-012-0061-5 (last access, September 27, 2019), 2012.

Apte, J. S., K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. H. Vermeulen, and S. P. Hamburg: High-resolution air pollution mapping with Google street view cars: exploiting big data, Environ. Sci. Tech., doi: 10.1021/acs.est.7b00891, 2018.

Baldauf, R., E. Thoma, M. Hays, R. Shores, J. Kinsey, B. Gullett, S. Kimbrough, V. Isakov, T. Long, R. Snow, A. Khlystov, J. Weinstein, F-L. Chen, R. Seila, D. Olson, I. Gilmour, S-H. Cho, N. Watkins, P. Rowley, and J. Bang: Traffic and meteorological impacts on near-road air quality: summary of methods and trends from the Raleigh near- road study, J. Air Waste Manage. Assoc. 58 (7), 865-878, doi: 10.3155/1047-3289.58.7.865, 2008.

Blanchard, C. L., E. L. Carr, J. F. Collins, T. B. Smith, D. E. Lehrman, H. M. Michaels: Spatial representativeness and scales of transport during the 1995 Integrated Monitoring Study in California's San Joaquin Valley, Atmos. Environ. 33, 4775-4786, https://doi.org/10.1016/S1352-2310(99)00284-8, 1999.

Boogaard, H., G. P. A. Kos, E. P. Weijers, N. A. H. Janssen, P. H. Fischer, S. C. van der Zee, J. J. de Hartog, and G. Hoek: Contrast in air pollution components between major streets and background locations: particulate matter mass, black carbon, elemental composition, nitrogen oxide and ultrafine particle number, Atmos. Environ. 45 (3), 650–658, doi.org/10.1016/j.atmosenv.2010.10.033, 2011.

Borrego, C., A.M. Costa, J. Ginja , M. Amorim, M. Coutinho, K. Karatzas, Th. Sioumis, N. Katsifarakis, K. Konstantinidis, S. De Vito, E. Esposito, P. Smith, N. André, P. Gérard, L.A. Francis, N. Castell, P. Schneider, M. Viana, M.C. Minguillón, W. Reimringer, R.P. Otjes, O. von Sicard, R. Pohle, B. Elen, D. Suriano, V. Pfister, M. Prato, S. Dipinto, and M. Penza: Assessment of air quality microsensors versus reference methods: the EuNetAir joint exercise, Atmos. Environ. 147, 246 – 263, http://dx.doi.org/10.1016/j.atmosenv.2016.09.050 (last access, February 22, 2019), 2016.

Brantley, H. L., G. S. W. Hagler, E. S. Kimbrough, R. W. Williams, S. Mukerjee, and L. M. Neas: Mobile air monitoring data-processing strategies and effects on spatial air pollution trends, Atmos. Meas. Tech. 7. 2169–2183. https://doi.org/10.5194/amt-7-2169-2014 (last access, September 27, 2019), 2014.

Caldwell, J. and Vahidsafa, A.: Propagation of Error, https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Supplemental_Modules_(Analytical_Chemistry)/Quantifying_Nature/Significant_Digits/Propagation_of_Error, 2019 (last access, March 23, 2020).

Camalier, L., S. Eberly, J. Miller, and M. Papp: Guideline on the Meaning and the Use of Precision and Bias Data Required by 40 CFR Part 58 Appendix A, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Air Quality Assessment Division, Air Quality Analysis Group, Research Triangle Park, North Carolina, https://www3.epa.gov/ttnamti1/files/ambient/monitorstrat/precursor/07workshopmeaning.pdf (last access, November 18, 2018), 2007.

Castell, N., F. R. Dauge, P. Schneider, M. Vogt, U. Lerner, B. Fishbain, D. Broday, and A. Bartonova: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, Environ. Int. 99, 293–302, http://dx.doi.org/10.1016/j.envint.2016.12.007 (last access, February 22, 2019), 2017.

Friberg, M. D., X. Zhai, H. A. Holmes, H. H. Chang, M. J. Strickland, S. E. Sarnat, P. E. Tolbert, A. G. Russell, and J. A. Mulholland: Method for fusing observational data and chemical transport model simulations to estimate spatiotemporally resolved ambient air pollution, Environ. Sci. Technol. 50 (7), 3695–3705, doi: 10.1021/acs.est.5b05134, 2016.

Gilbert, N.L., M.S., Goldberg, B., Beckerman, J.R., Brook, and M. Jerrett: Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model, J. Air Waste Manage. Assoc. 55, 1059-1063, 2005.

Goodman, L.: On the exact variance of products, J. American Statistical Assoc., December, 55 (292), 708-713, 1960. http://www.cs.cmu.edu/~cga/var/2281592.pdf (last access, March 23, 2020).

Google, Inc.: Table Details: California_2016_2017 https://console.cloud.google.com/bigquery?GK=street-view-air-quality&page=table&t=California_2016_2017&d=California_201605_201709_GoogleAclimaAQ&p=street-view-air-quality&redirect_from_classic=true&project=aclima-airview&folder=&organizationId= (last access, ~~October 7, 2019~~April 22, 2020),~~;~~ access permission granted by Google on request (https://goo.gl/EJMcCD), 2018.

Hankey, S. and J. D. Marshall: Land use regression models of on-road particulate air pollution (particle number, black carbon, PM2.5, particle size) using mobile monitoring, Environ. Sci. Technol. 49, 9194 – 9202, https://doi.org/10.1021/acs.est.5b01209, 2015.

Henderson, S.B., B. Beckerman, M. Jerrett, and M. Brauer: Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter, Environ. Sci. Technol. 41, 2422–2428, doi: 10.1021/es0606780. https://pubs.acs.org/doi/abs/10.1021/es0606780 (last access, February 21, 2019), 2007.

Herner, J. D., J. Aw, O. Gao, D. P. Chang, and M. J. Kleeman: Size and composition distribution of airborne particulate matter in northern California: I—particulate mass, carbon, and water-soluble ions, J. Air Waste Manage. Assoc. 55, 30–51, doi:10.1080/10473289.2005.10464600, 2005.

Holmes, H.A., X. Zhai, J. Redman, K. Digby, C. Ivey, S. Balachandran, S. A. Sororian, M. Friberg, W. Zhang, M. L. Maier, Y. Hu, A. G. Russell, J. A. Mulholland, H. H. Chang: Improved spatiotemporal source-based air pollutant mixture characterization for health studies, In: Steyn D., Mathur R. (eds) Air Pollution Modeling and its Application XXIII, Springer Proceedings in Complexity, Springer, Cham., https://doi.org/10.1007/978-3-319-04379-1_4 (last access, February 22, 2019), 2014.

Kanaroglou, P.S., M. Jerrett, J. Morrison, B. Beckerman, M.A. Arain, N.L. Gilbert, and J.R. Brook: Establishing an air pollution monitoring network for intraurban population exposure assessment: a location-allocation approach, Atmos. Environ. 39, 2399-2409, https://doi.org/10.1016/j.atmosenv.2004.06.049, 2005.

Karner, A. A., D. S. Eisinger, and D. A. Niemeier: Near-roadway air quality: Synthesizing the findings from real-world data: Environ. Sci. Technol. 44 (14), 5334–5344, doi: 10.1021/es100008x. https://pubs.acs.org/doi/10.1021/es100008x (last access, February 21, 2019), 2010.

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

Ku, H.: Notes on the use of propagation of error formulas, J. Research of National Bureau of Standards-C. Engineering and Instrumentation, 70C(4), 263-273, 1966, https://nvlpubs.nist.gov/nistpubs/jres/70C/jresv70Cn4p263_A1b.pdf (last access, March 23, 2020).

Laughner, J. L., Q. Zhu, and R. C. Cohen: The Berkeley high resolution tropospheric NO_2 product, Earth Syst. Sci. Data 10, 2069-2095, https://doi.org/10.5194/essd-10-2069-2018, 2018.

Lelieveld, J., J. S. Evans, M. Fnais, D. Giannadaki, D., A. Pozzer: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature 525 (7569), 367–371, https://www.nature.com/articles/nature15371 (last access, February 21, 2019), 2015.

Li, J. and P. Biswas: Optical characterization studies of a low-cost particle sensor, Aerosol and Air Quality Research 17, 1691–1704, doi: 10.4209/aaqr.2017.02.0085, http://aaqr.org/files/article/3435/1_AAQR-17-02-OA-0085_1691-1704.pdf (last access, February 22, 2019), 2017.

Lim, C. C., H. Kim, M.J. Ruzmyn Vilcassim, G. D. Thurston, T. Gordon, L.-C. Chen, K. Lee, M. Heimbinder, S.-Y. Kim: Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea, Environment International 131, 10502, https://doi.org/10.1016/j.envint.2019.105022 (last access, September 27, 2019), 2019.

Lunden, M. M. and B. LaFranchi: Metadata Documentation. https://docs.google.com/document/d/1qOR8j4ORSV9VhaAUVuLDnf9nfLh_RGbd7rCycXh3U3s/edit#heading=h.acka976x7igv (last access, November 14, 2018), 2017.

Marshall, J. D., E. Nethery, and M. Brauer: Within-urban variability in ambient air pollution: comparison of estimation methods, Atmos. Environ. 42 (6), 1359–1369, doi:10.1016/j.atmosenv.2007.08.012, 2008.

Messier, K. P., S. E. Chambliss, S. Gani, R. Alvarez, M. Brauer, J. J. Choi, S. P. Hamburg, J. Kerckhoffs, B. LaFranchi, M. M. Lunden, J. D. Marshall, C. J. Portier, A. Roy, A. A. Szpiro, R. C. H. Vermeulen, and J. S. Apte: Mapping air pollution with Google street view cars: efficient approaches with mobile monitoring and land use regression, Environ. Sci. Technol. 52, 12563 – 12572, doi: 10.1021/acs.est.8b03395, 2018.

Moore, D.K., M. Jerrett, W.J., Mack, and N. Kunzli: A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA, Journal of Environmental Monitoring 9, 246–252, doi:10.1039/b615795e, 2007.

Nam, E. K., T. E. Jensen, and T. J. Wallington. Methane emissions from vehicles, Environ. Sci. Technol. 38, 2005 – 2010, https://pubs.acs.org/doi/10.1021/es034837g (last access April 15, 2020), 2004.

Olson, D. A., D. M. Hammond, R. L. Seila, J. M. Burke, and G. A. Norris: Spatial gradients and source apportionment of volatile organic compounds near roadways, Atmos. Environ. 43 (35), 5647–5653, https://doi.org/10.1016/j.atmosenv.2009.07.065, 2009.

Ranasinghe, D., W.S. Choi, A.M. Winer and S.E. Paulson: Developing high spatial resolution concentration maps using mobile air quality measurements, Aerosol and Air Qual. Res. 16 (8), 1841-1853, doi:10.4209/aaqr.2015.07.0484, 2016.

Schneider, P., N. Castell, M. Vogt, F. R. Dauge, W. A. Lahoz, and A. Bartonova: Mapping urban air quality in near real-time using observations from low-cost sensors and model information, Environ. Int. 106, 234 – 247, https://doi.org/10.1016/j.envint.2017.05.005 (last access, February 22, 2019), 2017.

Seinfeld, J. and S. Pandis: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 3rd Edition, John Wiley and Sons, Inc., Hoboken, NJ, 2016.

Shusterman, A. A., J. Kim, K. J. Lieschke, C. Newman, P. J. Wooldridge, and R. C. Cohen: Observing local CO_2 sources using low-cost, near-surface urban monitors, Atmos. Chem. Phys. 18, 13773-13785, https://doi.org/10.5194/acp-18-13773-2018, 2018.

Simon, M. C., A. P. Patton, E. N. Naumova, J. I. Levy, P. Kumar, D. Brugge, and J. L. Durant: Combining measurements from mobile monitoring and a reference site to develop models of ambient ultrafine particle number concentration at residences, Environ. Sci. Technol. 52, 6985 – 6995, doi: 10.1021/acs.est.8b00292, 2018.

SCAQMD (South Coast Air Quality Management District): Site Survey Report for LAX -Hastings, http://www.aqmd.gov/docs/default-source/clean-air-plans/air-quality-monitoring-network-plan/aaqmnp-laxhastings.pdf?sfvrsn=16 (last access, February 28, 2019), 2018a.

SCAQMD (South Coast Air Quality Management District): Site Survey Report for Los Angeles (Central)-North Main Street, http://www.aqmd.gov/docs/default-source/clean-air-plans/air-quality-monitoring-network-plan/aaqmnp-losangeles.pdf?sfvrsn=16 (last access, February 28, 2019), 2018b.

SCAQMD (South Coast Air Quality Management District): Site Survey Report for Los Angeles-VA Hospital, http://www.aqmd.gov/docs/default-source/clean-air-plans/air-quality-monitoring-network-plan/aaqmnp-westlosangeles.pdf?sfvrsn=16 (last access, February 28, 2019), 2018c.

Stanaway, J. D. et al: Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392(10159), 1923-1994, doi: 10.1016/S0140-6736(18)32225-6, https://www.ncbi.nlm.nih.gov/pubmed/30496105 (last access July 26, 2019), 2018.

Steinle, S., S. Reis, and C. E. Sabel: Quantifying human exposure to air pollution—Moving from static monitoring to spatio-temporally resolved personal exposure assessment, Science of the Total Environment 443, 184–193, doi.org/10.1016/j.scitotenv.2012.10.098, 2013.

Stieb, D. M., S. Judek, and R. T. Burnett: Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season, J. Air Waste Manage. Assoc. 52, 470-484. doi: 10.1080/10473289.2002.10470794, 2002.

U.S. EPA: Risk and Exposure Assessment to Support the Review of the NO_2 Primary National Ambient Air Quality Standard, EPA-452/R-08-008a, https://www3.epa.gov/ttn/naaqs/standards/nox/data/20081121_NO2_REA_final.pdf (last access, June 24, 2019), 2008.

U.S. EPA: Quantitative Health Risk Assessment for Particulate Matter, EPA-452/R-10-005, https://www3.epa.gov/ttn/naaqs/standards/pm/data/PM_RA_FINAL_June_2010.pdf (last access, June 24, 2019), 2010a.

U.S. EPA: Quantitative Risk and Exposure Assessment for Carbon Monoxide – Amended, EPA-452/R-10-009, https://www3.epa.gov/ttn/naaqs/standards/co/data/CO-REA-Amended-July2010.pdf (last access, June 24, 2019), 2010b.

U.S. EPA: Health Risk and Exposure Assessment for Ozone, EPA-452/R-14-004a, https://nepis.epa.gov/Exe/ZyPDF.cgi/P100KBUF.PDF?Dockey=P100KBUF.PDF (last access, June 24, 2019), 2014.

U.S. EPA. Risk and Exposure Assessment for the Review of the Primary National Ambient Air Quality Standard for Sulfur Oxides, EPA-452/R-18-003, https://www.epa.gov/sites/production/files/2018-05/documents/primary_so2_naaqs_-_final_rea_-_may_2018.pdf (last access, June 24, 2019), 2018.

WHO: Air Quality Guidelines Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide, World Health Organization Regional Office for Europe, Denmark, http://www.euro.who.int/__data/assets/pdf_file/0005/78638/E90038.pdf (last access, September 27, 2019), 2006.

Zhang, K. M. and A. S. Wexler: Evolution of particle number distribution near roadways. Part I: analysis of aerosol dynamics and its implications for engine emission measurement, Atmos. Environ. 38, 6643-6653, https://doi.org/10.1016/j.atmosenv.2004.06.043 (last access, September 27, 2019), 2004.

Zhang, K. M., A. S. Wexler, Y. F. Zhu, W. C. Hinds, and C. Sioutas: Evolution of particle number distribution near roadways. Part II: the ‘road-to-ambient’ process, Atmos. Environ. 38, 6655–6665, https://doi.org/10.1016/j.atmosenv.2004.06.044 (last access, September 27, 2019), 2004.

Zhu, Y., W. C. Hinds, S. Kim, and C. Sioutas: Concentration and size distribution of ultrafine particles near a major highway, J. Air Waste Manage. Assoc. 52, 1032-1042, https://doi.org/10.1080/10473289.2002.10470842 (last access, September 27, 2019), 2002.

Table 1. Summary of driving dates and plans.

Location	Dates	Driving plan
San Francisco	May – Sept 2016 April – June 2017	Map every street in San Francisco, targeted driving
Los Angeles	Aug – Oct 2016	Map specific neighborhoods with repeat visits
San Joaquin Valley	Nov 2016 – Apr 2017 June – Sept 2017	Map multiple cities (Tracy, Stockton, Manteca, Merced, Modesto, Turlock), denser spatial coverage of Modesto

Table 2. Data sets used to evaluate spatial variability and to address individual research questions, including measurement uncertainty.

Location	Dates	Data analyses
San Francisco	May 1 – 31, 2017	Stationary vehicle collocated comparisons (side-by-side parking-garage car measurements); neighborhood spatial variability
Los Angeles	August 3 – 12, 2016	Stationary (side-by-side parking-garage) and moving vehicle collocated comparisons; neighborhood spatial variability; SCAQMD measurement audits
Los Angeles	September 20, 2016	Comparisons to stationary-monitor data; SCAQMD measurement audits
San Joaquin Valley	Nov 16 – 23, 2016	Stationary (side-by-side parking-garage) and moving vehicle collocated comparisons; urban-rural and interurban contrasts

Table 3. Sampling locations and dates for calibrations and audits through sample inlets conducted adjacent to stationary air quality monitors in Los Angeles.

Monitoring Site	Latitude	Longitude	Date
Long Beach near-road site (NRS) (W710)	33.86266	-118.19946	8/12/2016; 8/26/2016
Los Angeles International Airport (LAXH)	33.95500	-118.43028	9/20/2016
Santa Clarita	34.38342	-118.52822	10/6/2016; 10/25/2016

Table 4. Stationary monitoring sites in Los Angeles for which the SCAQMD provided high-resolution (1-minute) measurements. Hourly-average gas and $PM_{2.5}$ mass concentrations are available for other locations through EPA public data archives.

Code	Name	Latitude	Longitude	1-Minute Data	Scale[1]
CELA	Los Angeles N Main St[4]	34.0664	-118.2267	CO, NO, NO_2, O_3	Neighborhood
CMPT	Compton	33.9014	-118.2050	CO, NO, NO_2, O_3	Multiple[2]
HDSN	Long Beach (Hudson)	33.8022	-118.2197	CO, NO, NO_2, O_3	Neighborhood
LAXH	LAX-Hastings	33.9550	-118.4303	CO, NO, NO_2	Neighborhood
SLBH	South Long Beach[4]	33.7922	-118.1753		Neighborhood
W710	Long Beach Route 710	33.8594	-118.2003	$PM_{2.5}$ mass	Micro
WSLA	Los Angeles-VA Hospital	34.0508	-118.4564	CO, NO, NO_2, O_3	Multiple[3]

[1] EPA scales of representation are documented in Appendix D to Part 58 - Network Design Criteria for Ambient Air Quality Monitoring (https://www.law.cornell.edu/cfr/text/40/appendix-D_to_part_58, last access April 15, 2020). Neighborhood scale = 0.5 km to 4 km; middle scale = 100 m to 0.5 km; micro scale = several meters to ~100 m

[2] Neighborhood scale for O_3; middle scale for other species

[3] Middle scale for NO_2; neighborhood scale for O_3

[4] Hourly $PM_{2.5}$ or PM_{10} measurements available

Table 5. Performance summary of the gas-phase instruments (NO, NO_2, O_3, and CH_4) in parked vehicles (Lunden and LaFranchi, 2017).

Pollutant (Car)	Bias (ppbv)[1]	Precision (ppbv)[1]	Limit of Detection[2] (2σ, 1 sec) (ppbv)
NO (Coltrane)	± 2.1% + 0.3	± 2.3% ± 0.3	1.5
NO (Flora)	± 3.6% + 0.3	± 4.3% ± 0.3	1.7
NO_2 (Coltrane)	± 2.1% ± 0.4	± 2.8% ± 0.5	<0.1
NO_2 (Flora)	-2.4% + 0.2	± 2.2% ± 0.2	<0.1
O_3 (Coltrane)	± 2.1% ± 0.5	± 2.4% ± 0.6	1.8
O_3 (Flora)	± 2.0% ± 0.4	± 2.3% ± 0.5	1.8
CH_4 (Coltrane)	± 3.3	± 3.3	n/a

[1] Bias and precision are expressed as the upper bounds (at 90% confidence) of bias and precision metrics determined from differences between measured and target (audit) concentrations (Camalier et al., 2007).

[2] Limit of detection (LOD) is defined as the minimum concentration at which an observation can be discriminated from zero (with 95% confidence) at the specified sampling frequency (2 standard deviations of zero gas measurements).

Table 6. Performance of particle instruments (PN and BC) based on collocated parked vehicles. Evaluations performed between May 2016 and August 2017 (Lunden and LaFranchi, 2017).

Pollutant	Bias[1]	Precision[2]	RMSE[3]
$PN_{0.3-0.5}$	± 10.9%	± 9.8%	1293 c L^{-1} (1 sec)
			920 c L^{-1} (1 min)
$PN_{0.5-0.7}$	± 7.2%	± 7.5%	471 c L^{-1} (1 sec)
			237 c L^{-1} (1 min)
$PN_{0.7-1.0}$	± 11.3%	± 10.0%	170 c L^{-1} (1 sec)
			46 c L^{-1} (1 min)
$PN_{1.0-1.5}$	+ 25.7%	± 13.2%	69 c L^{-1} (1 sec)
			9 c L^{-1} (1 min)
$PN_{1.5-2.5}$	+ 25.7%	± 15.6%	71 c L^{-1} (1 sec)
			10 c L^{-1} (1 min)
BC	±11.9% ± 0.07 μg m^{-3}	not estimated	± 27.3% ± 0.26 μg m^{-3} (1 sec)
			± 15.6% ± 0.08 μg m^{-3} (10 sec)
			± 11.1% ± 0.05 μg m^{-3} (1 min)

[1] Bias for PN is calculated according to Camalier et al. (2007) where the values obtained by one car (Car A) are substituted for target (audit) concentrations. The positive sign of the bias estimate for the $PN_{1.0-2.5}$ (c L^{-1}) indicates a tendency of one instrument (Car B) to be biased high relative to the other instrument (Car A). Because BC concentrations were often close to LOD, bias for BC was estimated from linear least squares regression of bias vs concentration. A single bias value was estimated for each 6-hour collocation period using 1-minute aggregations from two vehicles. The bias estimates were regressed against the mean concentrations measured for the corresponding times. The relative and absolute components of bias were identified from the slope and intercept, respectively, of this linear regression ($r^2 = 0.37$, p-value < 0.0001).

[2] Precision is calculated according to Camalier et al. (2007) where the mean concentrations obtained by two cars are substituted for target (audit) concentrations.

[3] PN RMSE is determined from the vehicles' PN concentration differences relative to the means of the PN measured by the vehicles. RMSE for BC is estimated through a linear regression method (RMSE vs concentration) analogous to the procedure for estimating BC bias.

Table 7. Mean ambient concentrations and sample sizes as measured by the mobile platforms in each of the example study areas.[1]

Subset	NO (ppbv)	NO_2 (ppbv)	O_3 (ppbv)	CH_4 (ppmv)	$PN_{0.3-0.5}$ $(c\ L^{-1})^2$	$PN_{0.5-0.7}$ $(c\ L^{-1})^2$	$PN_{0.7-1.0}$ $(c\ L^{-1})^2$	$PN_{1.0-1.5}$ $(c\ L^{-1})^2$	$PN_{1.5-2.5}$ $(c\ L^{-1})^2$	$PN_{>2.5}$ $(c\ L^{-1})^2$
LA1[3]	10.5	15.3	44.1	NA	82,209	6725	1437	600	680	172
	589,555	626,136	228,498	NA	338,033	338,033	338,033	338,033	338,033	338,033
LA2[4]	21.4	22.5	37.7	2.17	42,818	4403	1251	537	748	274
	889,010	909,722	377,183	524,128	620,421	620,421	620,421	620,421	620,421	620,421
SJV1[5]	17.0	17.9	23.3	2.04	22,050	2769	742	304	375	153
	478,671	766,946	143,796	279,863	572,851	572,851	572,851	572,851	572,851	572,851
SJV2[6]	10.2	13.6	28.9	1.98	11,933	2527	1015	418	451	151
	294,514	393,917	35,215	140,022	283,179	283,179	283,179	283,179	283,179	283,179
SF[7]	6.0	10.3	26.5	1.98	13,947	4934	2288	922	868	154
	738,089	793,318	372,470	418,704	579,802	579,802	579,802	579,802	579,802	552,739

[1] Sample sizes are total number of 1-sec measurements summed across vehicles. Means are weighted by the number of measurements per vehicle.

[2] Particle number in size fractions 0.3 – 0.5 μm, 0.5 – 0.7 μm, 0.7 – 1.0 μm, 1.0 – 1.5 μm, 1.5 – 2.5 μm, > 2.5 μm.

[3] LA1 = Los Angeles, August 3 – 12, 2016 (8 days). BC, CH_4 = 1 car; NO, O_3, NO_2, and PN = 2 cars.

[4] LA2 = Los Angeles, September 12 – 30, 2016 (14 days). BC, CH_4 = 1 car; NO, O_3, NO_2, and PN = 2 cars.

[5] SJV1 = San Joaquin Valley, November 16 – 23, 2016 (6 days). BC, CH_4 = 1 car, NO and O_3 = 2 cars, NO_2 and PN = 3 cars.

[6] SJV2 = San Joaquin Valley, March 20 – 29, 2017 (6 days). BC, CH_4 = 1 car, NO and O_3 = 2 cars, NO_2 and PN = 2 cars.

[7] SF = San Francisco, May 1 – 12, 2017 (10 days). BC, CH_4 = 1 car; NO, O_3, NO_2, and PN = 2 cars.

Table 8. External calibration checks (zero and span) performed in Los Angeles with equipment and gas standards managed by the SCAQMD compared with internal checks performed by Aclima one month prior to the Los Angeles deployment, one month following this deployment, and during a 1-week return to San Francisco in the middle of the deployment. External and Aclima calibration checks were conducted through the inlet lines of the mobile platforms.

Species	Audit	Bias (% ± ppbv)	Precision (% ± ppbv)	Number of Span Checks	Number of Zero Checks
NO	Aclima	± 3.5% + (< 1)	± 4.5% + (< 1)	22	22
	SCAQMD	± 8.2% + (< 1)	± 6.0% + (< 1)	10	10
NO_2	Aclima	- 3.7% ± 0.4[1]	± 3.7% ± 0.4	19	20
	SCAQMD	- 1.9% ± 0.6[1]	± 4.9% ± 0.6	6	10
O_3	Aclima	± 2.4% ± 0.9	± 2.3% ± 1.1	20	18
	SCAQMD	± 3.3% ± 1.2	± 3.8% ± 1.5	10	10

[1] Negative bias only

Table 9. Performance summary for measurements reported by collocated vehicles (mean difference ± 1 standard deviation; mean concentrations in parentheses). Standard deviations are reported here to indicate the variability of the 1-s differences. Mean differences provide a measure of average intervehicle differences. For periods when three vehicles were driven, the largest mean difference between vehicles is listed. The signs of the mean differences are not indicated because no vehicle is an audit standard. All values were determined from 1-s time resolution data.

Setting	Period[1]	NO[2] (ppbv)	O_3[2] (ppbv)	NO_2[2] (ppbv)	$PN_{0.3-0.5}$[2] ($c\ L^{-1}$)
Parking ~~structure[3]~~lot[3]	LA1	0.6 ± 49.5 (11.3)	1.5 ± 8.1 (41.6)	0.3 ± 12.0 (15.8)	18346 ± 21024 (81929)
Parking ~~structure[3]~~lot[3]	LA2	3.9 ± 66.9 (21.5)	1.0 ± 9.9 (34.6)	1.9 ± 14.3 (22.7)	6525 ± 20049 (44058)
Parking structure[3]	SJV1	0.5 ± 2.1 (3.8)	4.5 ± 2.3 (18.9)	1.0 ± 1.5 (16.7)	1126 ± 3922 (12527)
Parking structure[3]	SF	0.2 ± 7.5 (3.5)	0.8 ± 3.9 (24.2)	1.1 ± 5.7 (6.2)	507 ± 1865 (14154)
Moving[4], < 10 m	SJV1[8]	NA[10]	NA[10]	5.6 ± 32.7 (15.1)	132 ± 4242 (7661)
Moving[5], 10–100 m	SJV1[8]	NA[10]	NA[10]	1.9 ± 20.1 (16.2)	454 ± 2478 (5883)
Moving[6], < 10 m	SF-LA[9]	13.8 ± 56.7 (27.9)	1.8 ± 2.2 (40.9)	3.4 ± 9.4 (16.8)	20797 ± 5410 (64187)
Moving[7], 10–100 m	SF-LA[9]	5.1 ± 49.1 (26.5)	0.5 ± 3.2 (42.2)	1.0 ± 12.1 (17.3)	19294 ± 7670 (60046)

[1] LA1 = August 3 – 12, 2016 (8 days); LA2 = September 12 – 30, 2016 (14 days); SJV1 = November 16 – 23, 2016 (6 days); SJV2 = March 21 – 30, 2017 (6 days)

[2] Vehicle-to-vehicle concentration differences were determined from 1-s measurements. Means and standard deviations of paired differences were determined for each data pair. Time periods when a vehicle was sampling through a calibration port (whether a calibration was in process) were excluded to ensure that vehicles were sampling the same ambient air for all comparisons.

[3] ~~One~~ The parking ~~structure is~~lot in Los Angeles ~~and~~ was used for LA1 and LA2. The ~~second~~ parking structure ~~is~~ in San Francisco ~~and~~ was used for all SF and SJV drives.

[4] intervehicle distance < 10 m (average = 5 m), average speed = 3.0 m s^{-1} (10.6 km h^{-1})

[5] intervehicle distance 10 – 100 m (average = 32 m), average speed = 25.6 m s^{-1} (92.0 km h^{-1})

[6] intervehicle distance < 10 m (average = 6 m), average speed = 5.9 m s^{-1} (21.2 km h^{-1})

[7] intervehicle distance 10 – 100 m (average = 44 m), average speed = 27.7 m s^{-1} (99.7 km h^{-1})

[8] November 16, 2016 (I-580 and other locations, Flora and Rhodes, Figure S2)

[9] August 1, 2016 driving from San Francisco to Los Angeles (I-5 and other locations)

[10] Not available. SJV1, one car (Rhodes) of collocated moving pair lacked NO and O_3 samplers.

Table 10. Comparison of mobile-platform to collocated stationary-site measurements made at LAXH on September 20, 2016. The two cars alternated positions between an audit location 6.6 m for Coltrane and 8.5 m for Flora horizontal distance from the ground-level coordinates of the LAXH monitor (inlet situated 4.2 m agl inside a fenced enclosure) and a sampling location further from the monitor (24.1 m for Coltrane and 18.5 m for Flora). Data from the audit tests are excluded. The Coltrane audit period was 10:22 a.m. – 12:20 p.m. PDT (n = 119). The Flora audit period was 9:19 a.m. – 10:20 p.m. PDT (n = 56). The means ± standard errors of the means were determined for each car from the 1-minute measurements made at the two distances from the stationary monitor. Standard errors indicate the uncertainties of the mean concentrations and mean differences. Differences of 1-minute measurements were determined prior to averaging. The variabilities of the 1-minute differences can be obtained by multiplying standard errors by square root of sample size (n).

Platform	N[1]	NO (ppbv)	NO_2 (ppbv)	O_3 (ppbv)	O_x (ppbv)[2]
Coltrane	56	17.7 ± 1.1	37.7 ± 1.8	16.2 ± 0.6	53.9 ± 1.3
Flora	56	18.8 ± 1.2	37.0 ± 1.7	ND	ND
LAXH	56	19.0 ± 1.2	36.6 ± 1.7	20.3 ± 0.4	56.9 ± 1.3
Coltrane - LAXH	56	-1.3 ± 0.4	1.2 ± 0.4	-4.1 ± 0.3	-3.0 ± 0.2
Flora - LAXH	56	-0.2 ± 0.3	0.6 ± 0.3	ND	ND
Coltrane	119	4.5 ± 0.4	16.5 ± 1.0	41.7 ± 0.4	52.2 ± 0.6
Flora	119	3.1 ± 0.3	14.7 ± 0.9	36.1 ± 0.7	50.7 ± 0.3
LAXH	119	4.1 ± 0.4	14.6 ± 0.9	38.8 ± 0.7	53.4 ± 0.3
Coltrane - LAXH	119	-0.1 ± 0.2	0.3 ± 0.2	-0.9 ± 0.3	-0.2 ± 0.3
Flora - LAXH	119	-1.0 ± 0.1	0.04 ± 0.2	-2.6 ± 0.2	-2.6 ± 0.2

[1] Total minutes. Flora audit period 9:19 a.m. – 10:20 p.m. PDT (n = 56) and Coltrane audit period 10:22 a.m. – 12:20 p.m. PDT (n = 119). Sample sizes for individual measurements may be smaller due to excluding audit values. Mean paired differences are computed only for non-audit samples.

[2] $O_x = NO_2 + O_3$

Table 11. Dates, locations, and times when vehicle pairs sampled different areas within the northern San Joaquin Valley.

Date	Areas Sampled	Vehicles	Hours	Mean Distance (km)	Species Measured by Both Vehicles
Nov 16	Stockton – Rural	Flora – Coltrane	12 – 14	56.2	NO NO_2 O_3 PM
Nov 16	Stockton – Tracy	Flora – Rhodes	12 – 14	37.5	NO_2 PM
Nov 17	Stockton – Manteca	Coltrane – Flora	13 – 14	18.7	NO NO_2 O_3 PM
Nov 17	Stockton – Stockton	Coltrane – Rhodes	13 – 14	1.2	NO_2 PM
Nov 17	Stockton – Manteca	Rhodes – Flora	12 – 15	17.9	NO_2 PM
Nov 18	East – West SJV	Flora – Coltrane	12 – 14	49.9	NO NO_2 O_3 PM
Nov 18	East – West SJV	Rhodes – Coltrane	12 – 14	49.7	NO_2 PM
Nov 21	East – West SJV	Flora – Rhodes	12 – 14	47.1	NO_2 PM
Nov 21	East – West SJV	Coltrane – Rhodes	12 – 14	37.9	NO_2 PM
Nov 22	Stockton – Modesto	Flora – Coltrane	12 – 14	43.9	NO NO_2 O_3 PM
Nov 22	Stockton – Modesto	Flora – Rhodes	12 – 14	44.6	NO_2 PM
Nov 23	Stockton – Modesto	Flora – Rhodes	10 – 13	30.4	NO_2 PM
Nov 23	Stockton – Highway	Flora – Coltrane	10 – 13	61.0	NO NO_2 O_3 PM

Table 12. Fractional mean differences (FMD) when vehicle pairs sampled different areas within the northern San Joaquin Valley. Vehicles A and B correspond to the first and second areas sampled, respectively. Uncertainties are one standard error of the means. NA = not available; one car (Rhodes, R) measured only NO_2 and PM concentrations.

Date	Areas Sampled	Car[1] A–B	NO ~~FMD~~	NO_2 ~~FMD~~	O_3 ~~FMD~~	PM ~~FMD~~
Nov 16	Stockton – Rural	F–C	-0.30 ± 0.02	0.30 ± 0.02	-0.24 ± 0.004	0.96 ± 0.01
Nov 16	Stockton – Tracy	F–R	NA	0.46 ± 0.02	NA	0.14 ± 0.01
Nov 17	Stockton – Manteca	C–F	0.61 ± 0.03	-0.16 ± 0.01	0.01 ± 0.004	0.02 ± 0.004
Nov 17	Stockton – Stockton	C–R	NA	0.007 ± 0.01	NA	0.11 ± 0.003
Nov 17	Stockton – Manteca	R–F	NA	-0.18 ± 0.01	NA	0.12 ± 0.004
Nov 18	East – West SJV	F–C	-0.61 ± 0.05	-0.30 ± 0.02	NA	0.23 ± 0.004
Nov 18	East – West SJV	R–C	NA	-0.23 ± 0.02	NA	0.14 ± 0.004
Nov 21	East – West SJV	F–R	NA	-0.30 ± 0.02	NA	-0.13 ± 0.008
Nov 21	East – West SJV	C–R	NA	0.30 ± 0.02	NA	0.23 ± 0.006
Nov 22	Stockton – Modesto	F–C	0.36 ± 0.03	0.49 ± 0.01	-0.42 ± 0.01	0.10 ± 0.005
Nov 22	Stockton – Modesto	F–R	NA	0.70 ± 0.01	NA	-0.12 ± 0.006
Nov 23	Stockton – Modesto	F–R	NA	-0.09 ± 0.01	NA	-0.65 ± 0.02
Nov 23	Stockton – Highway	F–C	0.40 ± 0.03	-0.09 ± 0.01	0.18 ± 0.006	-0.57 ± 0.02

[1] C = Coltrane, F = Flora, R = Rhodes

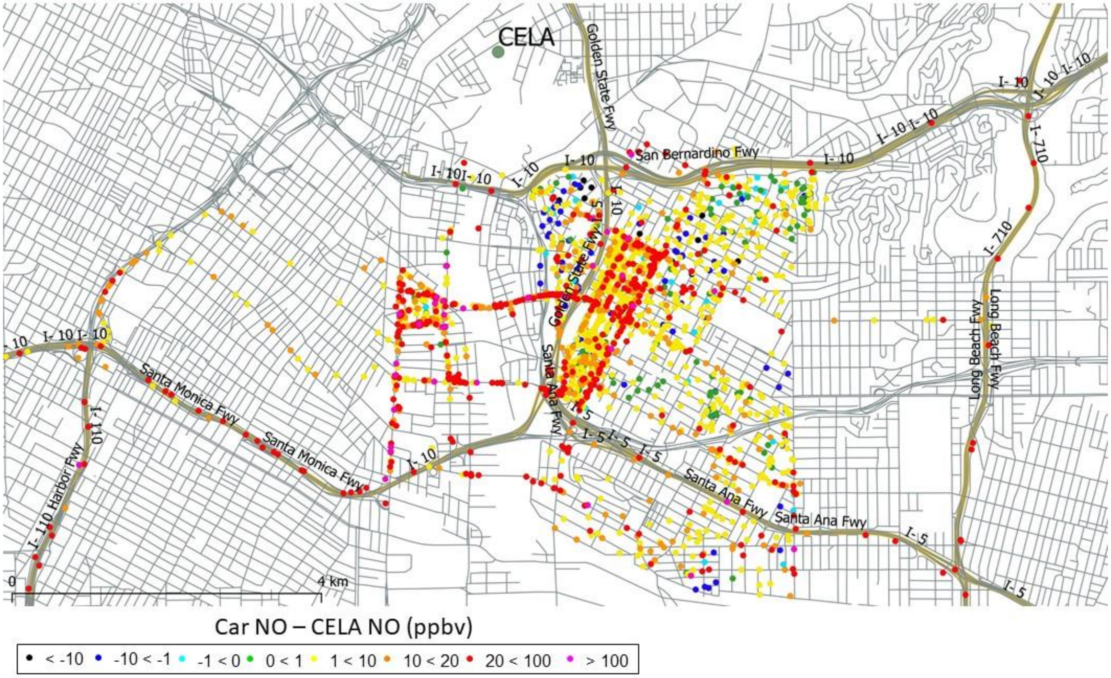


Figure 1. Paired differences in 1-minute NO concentrations measured by cars and by the air quality monitor in downtown Los Angeles (CELA) during August 3 – 12, 2016. Map generated with QGIS version 3.2.2 (https://qgis.org/en/site/) open-source software licensed under the GNU General Public License (http://www.gnu.org.licenses). California state highway shapefiles obtained from the OpenStreetMap community (www.openstreetmap.org) and MapCruzin (www.mapcruzin.com), licensed under the Creative Commons Attribution Share-Alike 2.0 license. U.S. highways shapefile obtained from U.S. Bureau of the Census TIGER/Line shapefiles public data (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html).

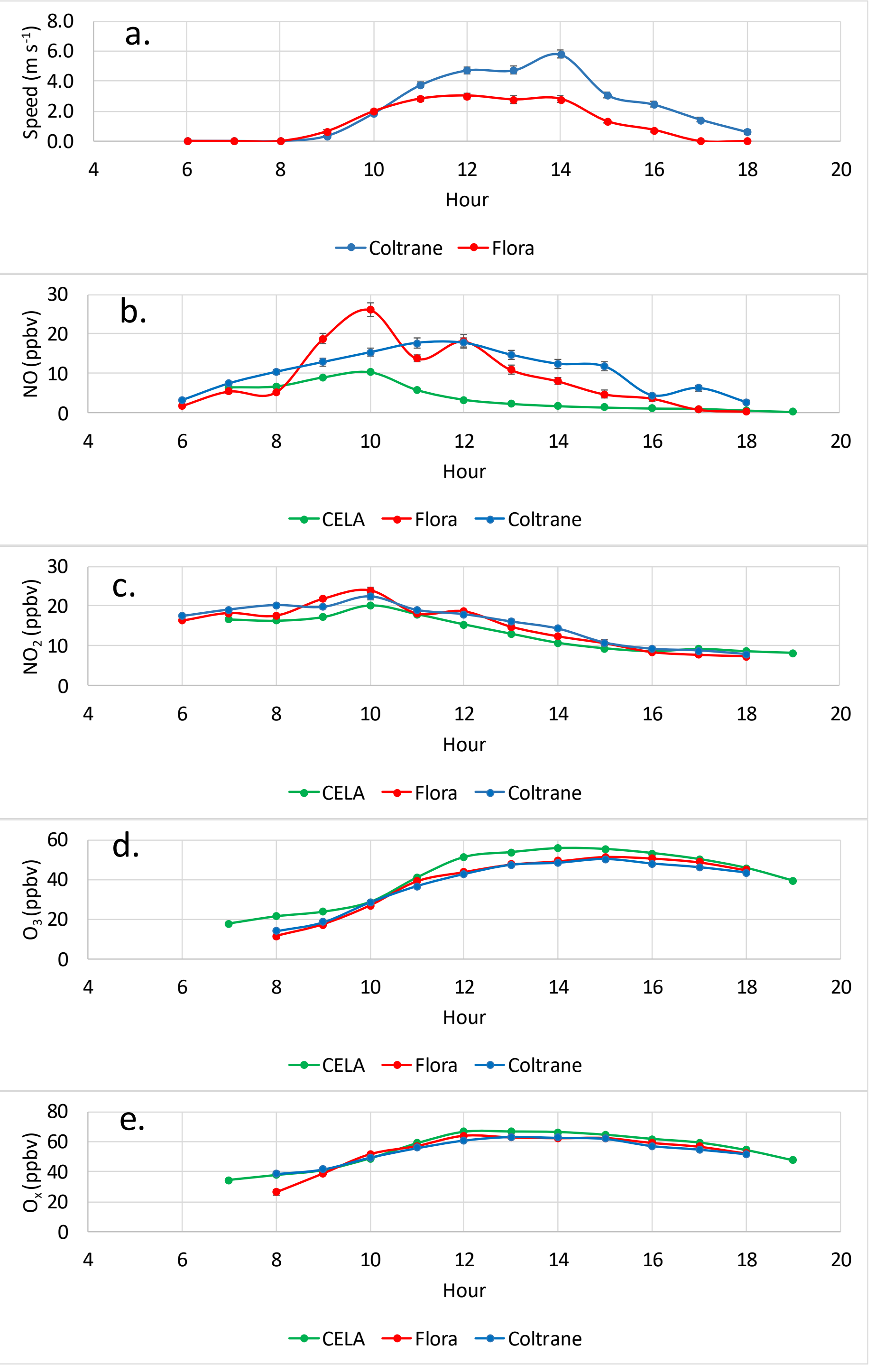


Figure 2~~1~~. Mean vehicle speeds and pollutant concentrations averaged by hour over all Los Angeles driving days between August 3 and 12, 2016. Standard errors of the means are plotted but are generally smaller than the symbol sizes.

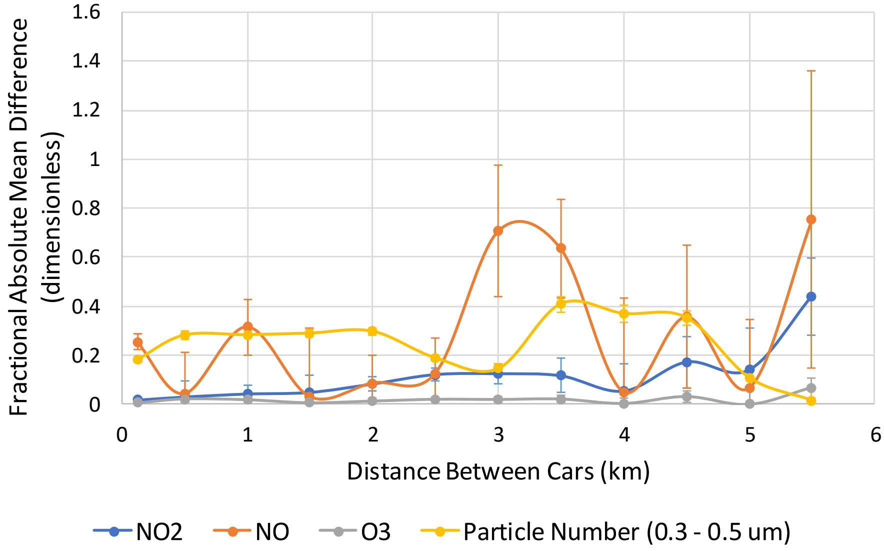


Figure 3~~2~~. Intervehicle FAMD vs mean intervehicle distance associated with sampling in Los Angeles (near CELA) from August 3 – 12, averaged over 0.5 km bins (0 – 0.25 km, 0.25 – 0.75 km, etc.). Error bars are 1-sigma uncertainties determined as described in the definition of FAMD. The sizes of the error bars reflect variations in the number of samples in each bin (N = 14 to 2433) as well as sampling variability.

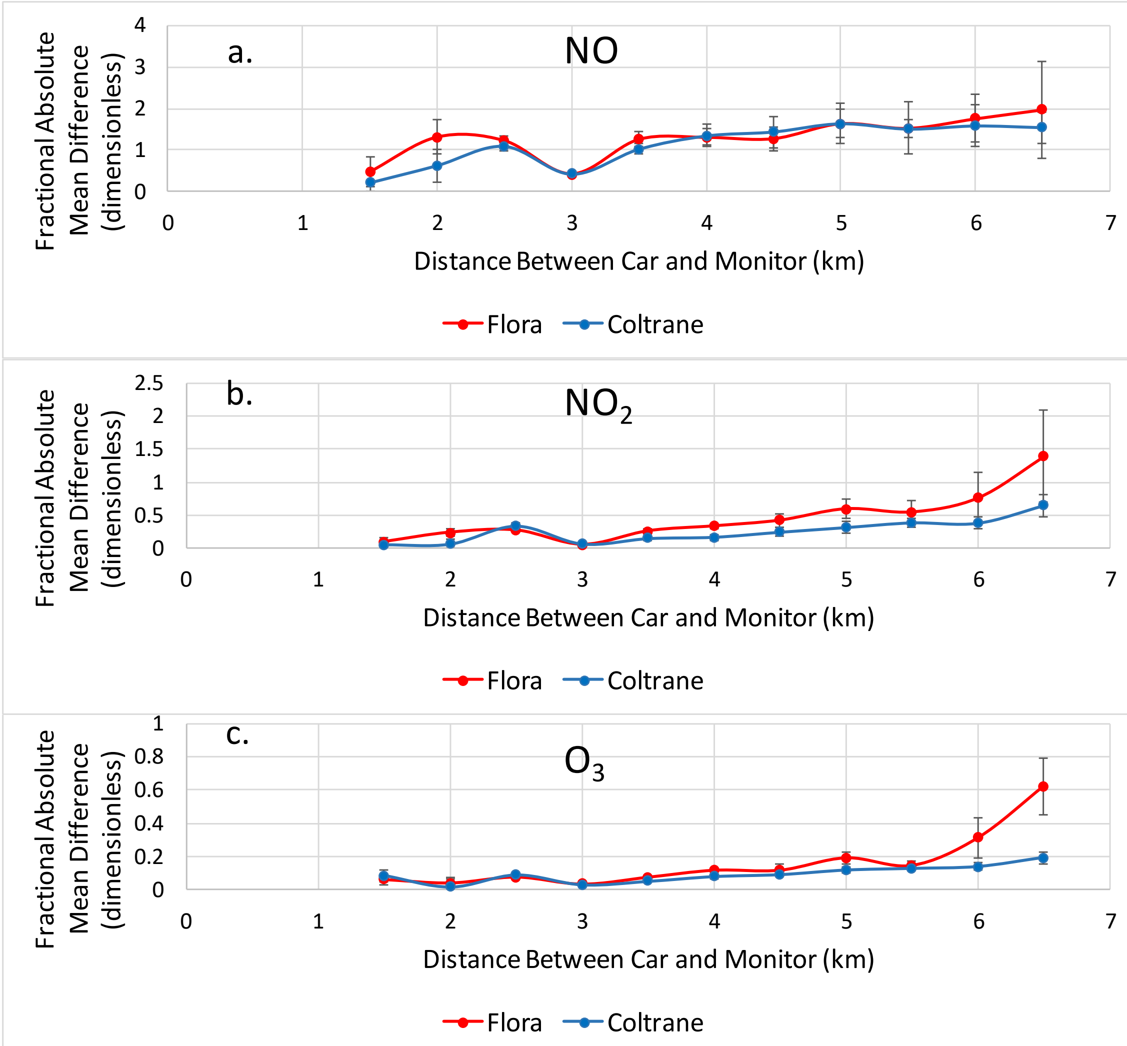


Figure 4~~3~~. Fractional absolute mean difference (FAMD) for (a) NO, (b) NO_2, and (c) O_3 vs mean ~~-~~intervehicle distance for August 3 – 12, 2016, Los Angeles sampling, averaged over 0.5 km bins. Error bars are 1 sigma uncertainties as described in the text. The sizes of the error bars reflect variations in the number of samples in each bin (N = 3 - 19 at 6.5 km to 222 – 338 at 3.5 km). The 3 km bin (N = 1273 – 3906) consists primarily of measurements made in the parking garage.

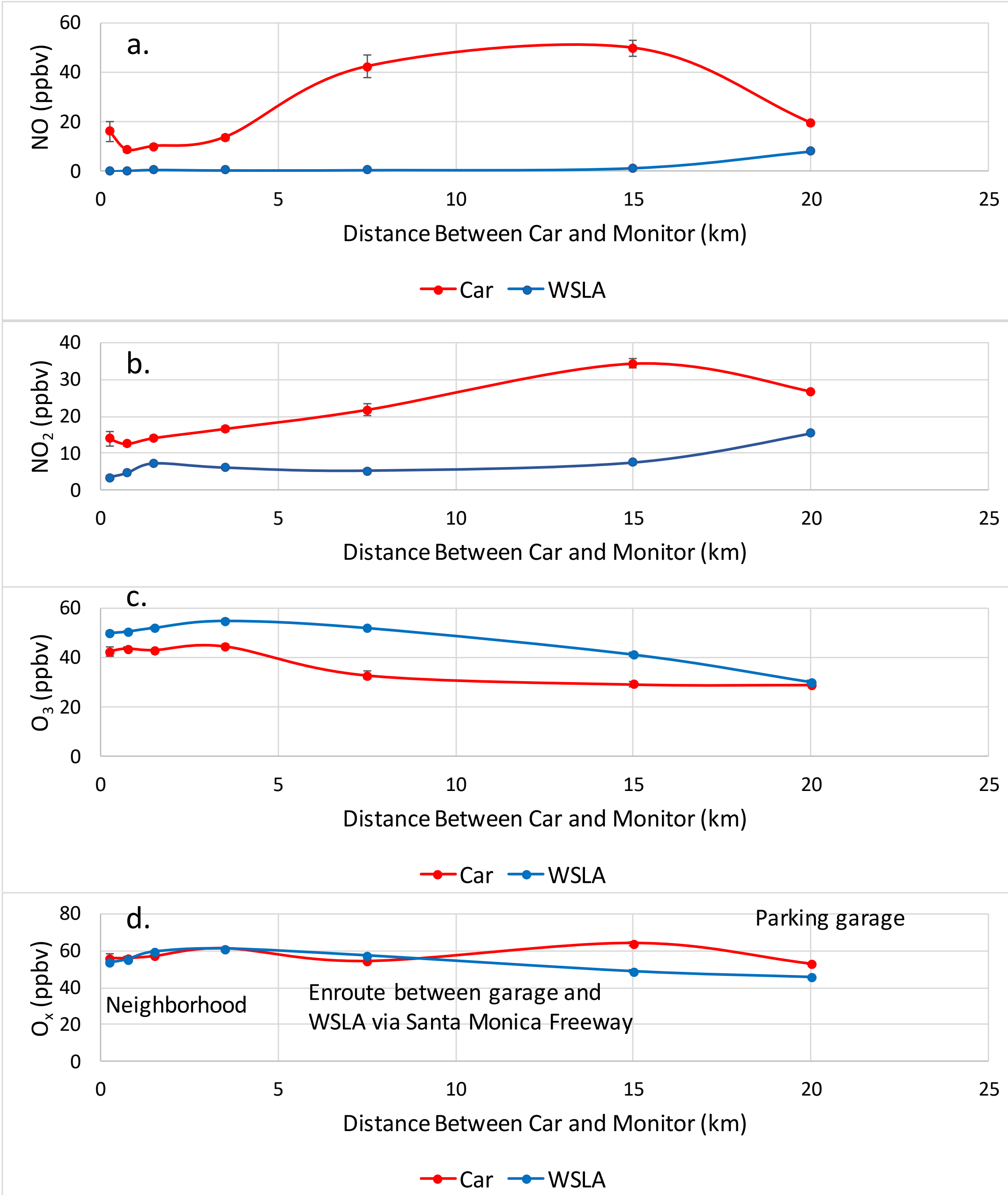


Figure 54. Mobile platform monitoring and WSLA measurements versus distance between cars and WSLA on four days (September 13, 19, 26, and 29, 2016) when the cars drove near WSLA. The first bin includes all distances less than 0.5 km; the minimum distance between cars and monitor was 158 m. Locations are indicated. Standard errors of the means are shown but most are smaller than the symbols.

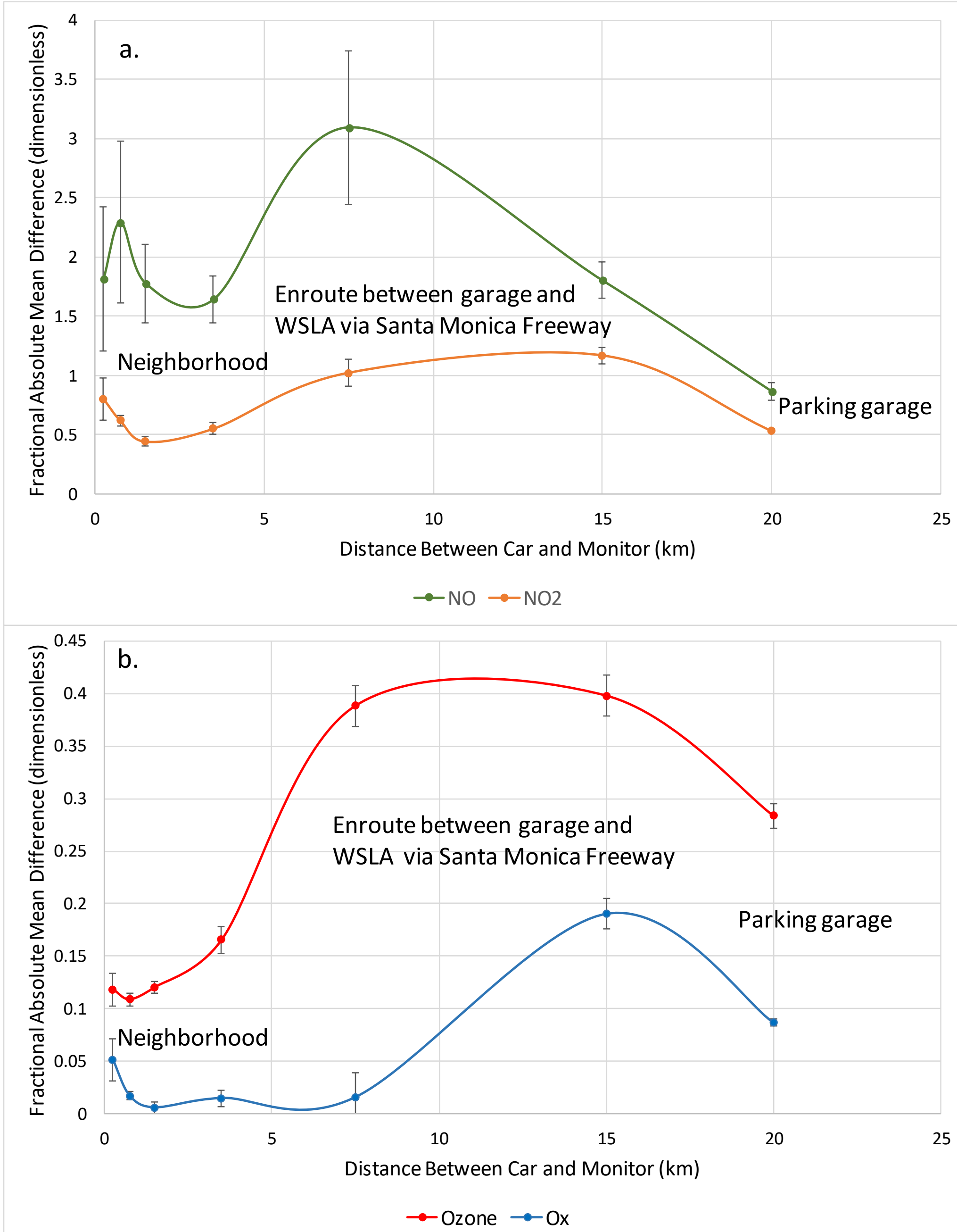
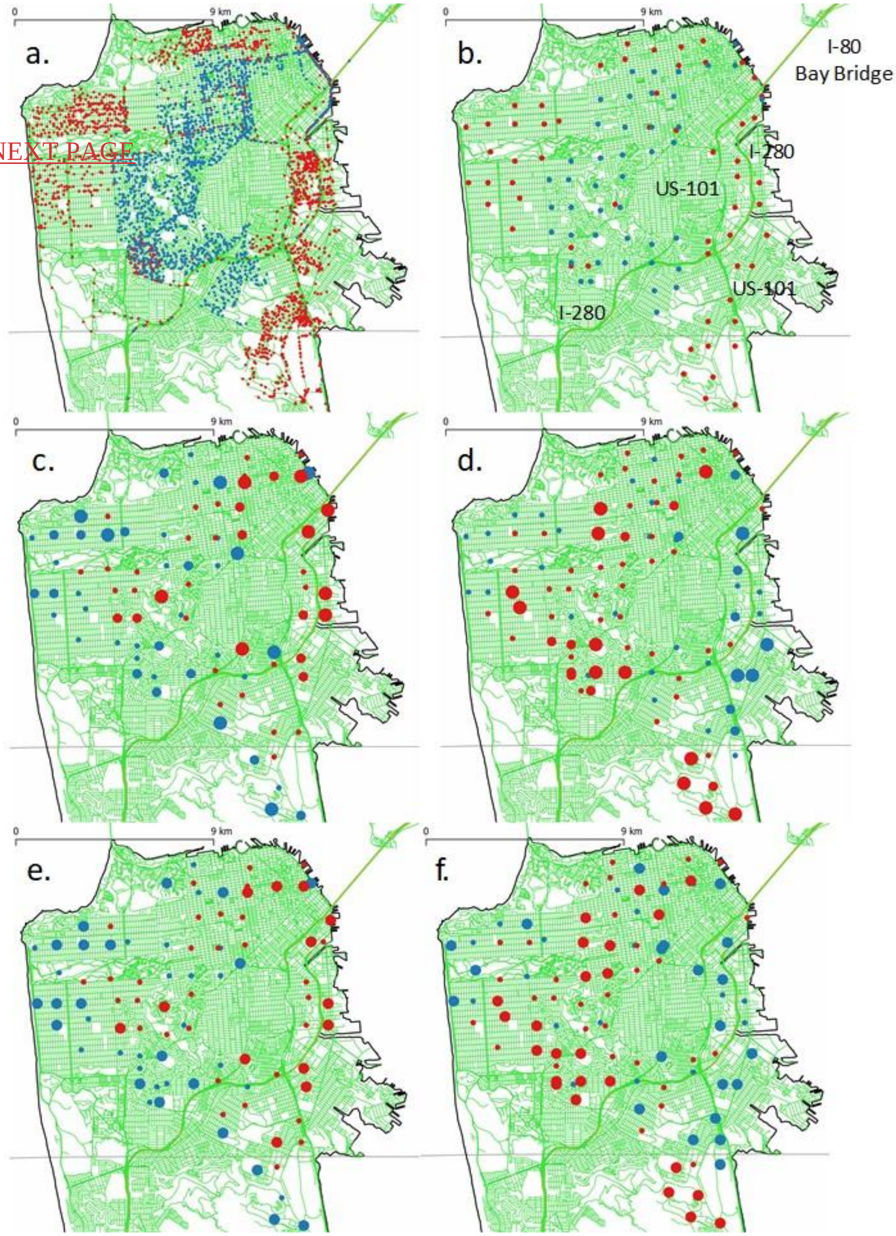


Figure 6~~5~~. FAMD between mobile platform monitoring and WSLA measurements versus distance between cars and WSLA on four days (September 13, 19, 26, and 29, 2016) when the cars drove near WSLA. The first bin includes all distances less than 0.5 km; the minimum distance between cars and monitor was 158 m. Locations are indicated. One-sigma uncertainties of the FAMD were determined as described in the definition of FAMD in the text.

REPLACE WITH NEXT PAGE



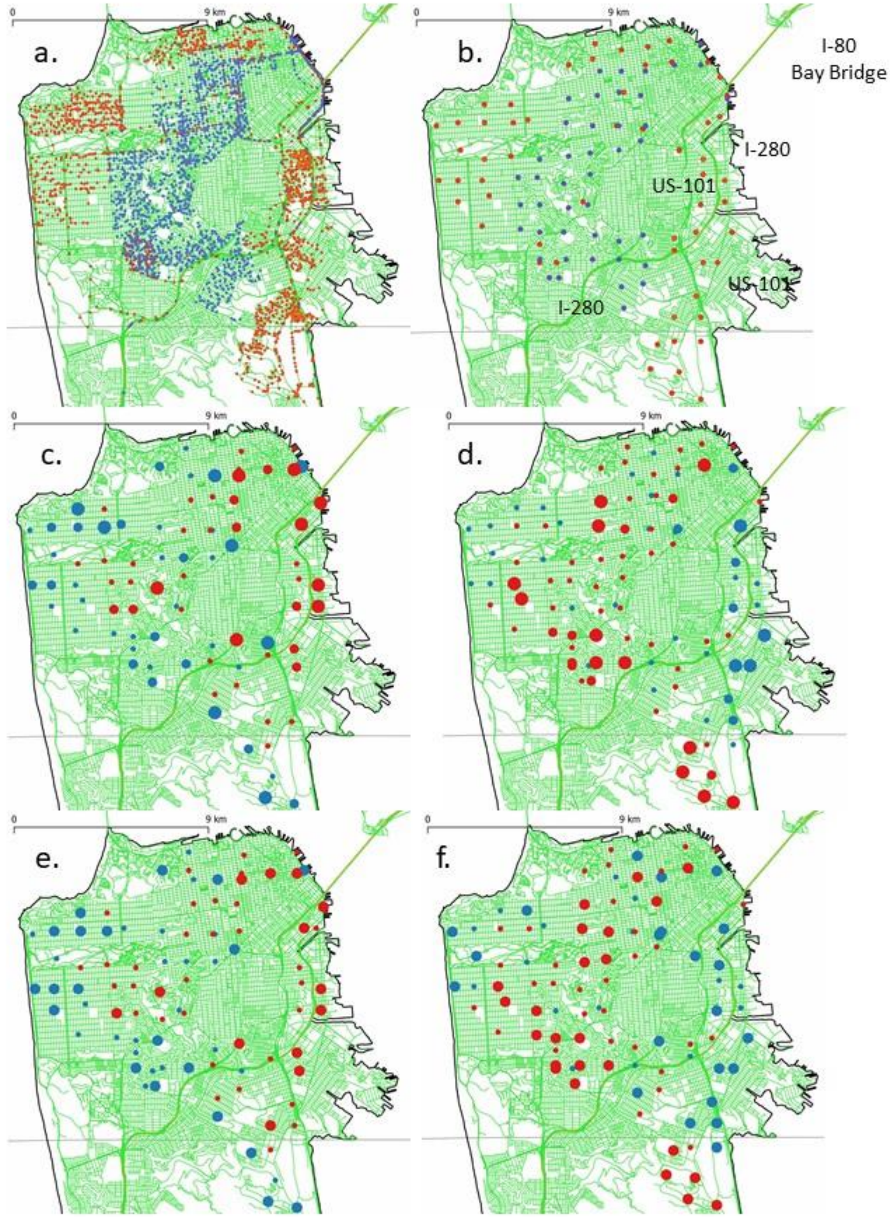


Figure 76. San Francisco sampling locations and results for May 1 through 12, 2017: (a) 1-minute resolution locations (~~red~~ gold = Flora, ~~blue~~ lavender = Coltrane), (b) 1-kilometer resolution locations (~~red~~ gold = Flora, ~~blue~~ lavender = Coltrane), (c) NO_2 intervehicle differences (red = positive, blue = negative; ~~small~~ large symbol = < -4 or > 4 ppbv, medium = -4 to -2 or 2 to 4 ppbv, ~~large~~ small = -2 to +2 ppbv), (d) O_3 intervehicle differences (same scale as NO_2), (e) NO_2 FMD (red = positive, blue = negative; ~~small~~ large symbol = < -0.5 or > 0.5, ~~large~~ small = -0.5 to +0.5), (f) O_3 FMD (red = positive, blue = negative; ~~small~~ large = < -0.05 or > 0.05, ~~large~~ small = -0.05 to +0.05). Maps generated with QGIS version 3.2.2 (https://qgis.org/en/site/) open-source software licensed under the GNU General Public License (http://www.gnu.org.licenses). California coastline shapefile obtained from the OpenStreetMap community (www.openstreetmap.org) and MapCruzin (www.mapcruzin.com), licensed under the Creative Commons Attribution Share-Alike 2.0 license. U.S. highways and California county boundary shapefiles obtained from U.S. Bureau of the Census TIGER/Line shapefiles public data (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html).

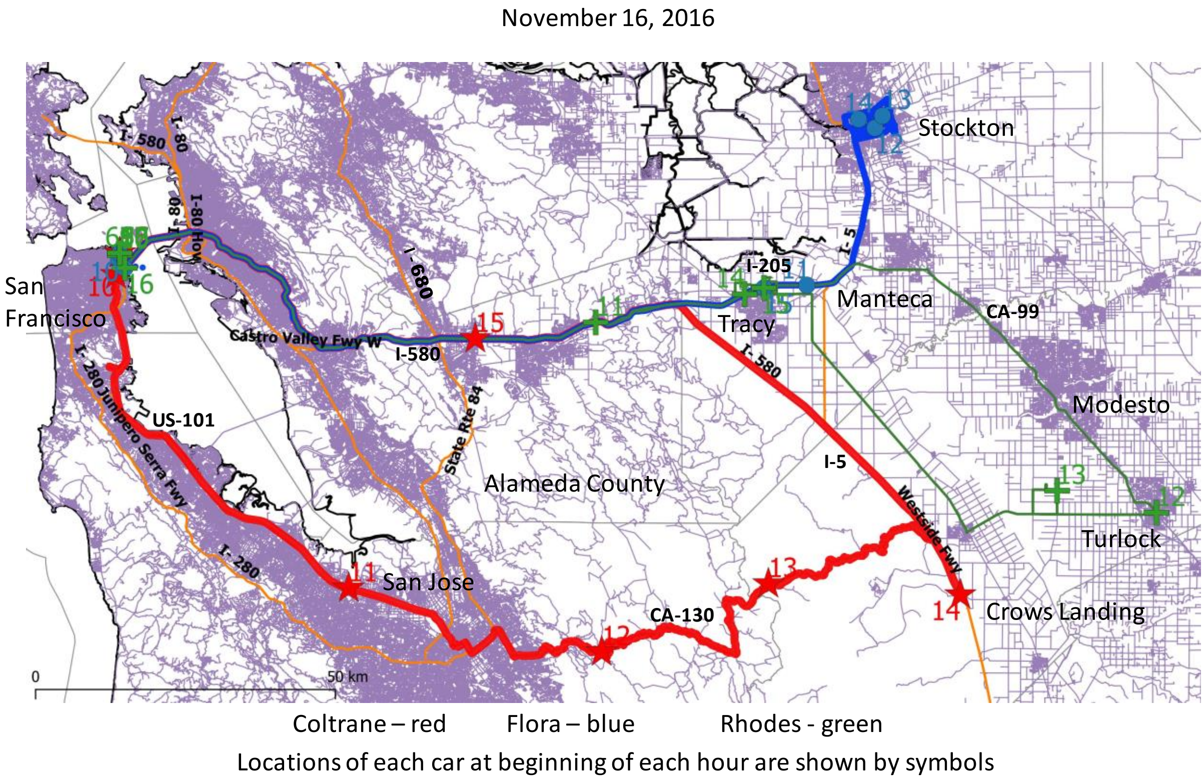


Figure 8~~7~~. San Joaquin Valley driving routes on Nov 16, 2016. The positions of each car at the beginning of each hour are marked. The drives began and ended at the parking garage in San Francisco. Locations of cities identified in the text are also shown. Map generated with QGIS version 3.2.2 (https://qgis.org/en/site/) open-source software licensed under the GNU General Public License (http://www.gnu.org.licenses). California coastline and state highway shapefiles obtained from the OpenStreetMap community (www.openstreetmap.org) and MapCruzin (www.mapcruzin.com), licensed under the Creative Commons Attribution Share-Alike 2.0 license. U.S. highways and California county boundary shapefiles obtained from U.S. Bureau of the Census TIGER/Line shapefiles public data (https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html).