Community Air Monitoring Plan: Appendix D

California Statewide Mobile Monitoring Initiative (SMMI) Hyperlocal Ambient Concentration Estimate Validation and Quality Assurance System (v2.2)

aclima

July 1, 2025

CALIFORNIA AIR RESOURCES BOARD

CALIFORNIA CLIMATE INVESTMENTS
Cap and Trade Dollars at Work

The Statewide Mobile Monitoring Initiative is part of California Climate Investments, a statewide initiative that puts billions of Cap-and-Trade dollars to work reducing greenhouse gas emissions, strengthening the economy, and improving public health and the environment — particularly in disadvantaged communities.

Table of Contents

1.0 Introduction

Among the useful data outputs that mobile mapping generates are analyses that support visualization of typical pollution concentrations with high spatial resolution (“ Hyperlocal Maps”) across a specific region, city, or community. These maps highlight typical concentrations over a defined measurement period illustrating high and low pollution concentrations at the street level. The maps are only one of many data products that can be generated from mobile mapping, but represent a foundational data product that fills a critical gap in understanding the spatial distribution of pollution. Aclima uses verified 1-Hz data to produce concentration estimates at desired and practical aggregation length scales (e.g. hexbins or road segments) showing areas of persistently high or low levels of individual pollutants, supporting the identification of areas of disproportionate impact..

Aclima's Data Quality Objectives for the creation of maps of ambient concentration estimates are as follows:

- Produce ambient pollution concentrations estimates for the monitoring time period and monitoring area from measurements collected at different times of day and throughout the week and across seasons to adequately address seasonal and diurnal variations in the data.
- Spatial distribution of data throughout the entire user-defined geographic area.
- Produce concentration estimates at desired and practical spatial aggregation scales (e.g. hexbins, road segments).
- Include a measure of confidence (i.e. confidence interval) with each ambient pollution concentration estimate in order that users can understand the reliability of the estimates and assess the true difference in concentrations between locations.
- Monitor and track the performance of each pollution measurement using key data quality indicators of bias, precision, and drift.

We support the creation of ambient concentration estimates for the following pollutants[1]:

- O_3, NO_2, CO, CO_2, $PM_{2.5}$, and BC

Ambient pollution concentration estimates are not produced until data collection is complete and the data are verified. These concentration estimates are calculated using a data-driven modeling framework that uses all the verified data on that road or in that area and incorporates additional sources of data that provide information on broad regional spatial trends and region-wide temporal trends, including, in some cases, regulatory measurements from stationary monitoring sites. The results provide the best estimate of the average concentration for each pollutant that retains true spatial variability at the hyperlocal scale while accounting for biases that may result from the mobile method sampling at different times in different locations. This high spatial resolution of Aclima's resulting data products supports identification of emissions sources as well as information on neighborhood-level variability in air pollution concentrations to support disparity analysis. The mobile mapping method is not a reference method designed to support the National Ambient Air Quality Standards (NAAQS), which are

[1] The TVOC sensor has two characteristics that make data from the sensor unsuitable to support ambient concentration estimates; (1) the sensor is sensitive to a wide range of VOCs with the sensitivity to different classes of VOCs varying by multiple orders of magnitude, (2) the sensor is prone to baseline drift. For more information, see Appendix C, Section 5.6.5.

supported by a network of stationary reference monitors. Thus, data products from the mobile method do not support assessment of compliance with NAAQS.

The Data Quality Objectives are used to define the sampling, measurement, data processing, and analysis methodologies. This document describes the process Aclima follows moving from the verified 1-Hz data to high-resolution spatial estimates of ambient concentration, including how we test for data that might not be valid for inclusion in calculations or visualizations as well as our validation and verification processes.

2.0 Data preparation

2.1 Data geolocation and aggregation

As Aclima's cars drive along publicly accessible roads, sensors within the Aclima Mobile Note ("AMN") sample at a 1-second frequency. These measurements are associated in the Aclima database with a specific 1-s Global Positioning System (GPS) time and location. The raw GPS position information can at times be some meters from the road the car was driving due to the fundamental uncertainty in the GPS measurement as well as external factors, such as tall buildings, interfering with the ability of the GPS system to achieve a solid location fix. The position of the raw GPS data is corrected to align with the route driven by the car, often termed "snapping to the road", reducing location uncertainty (Figure 1).

The 1-second measurements are assigned to a unique spatial unit (i.e. road segment, hexbin, etc.) that supports a specific monitoring objective based on the corrected location (latitude and longitude) of the data point. Each individual drive in that spatial unit is defined as a pass or visit. First the 1-second data is aggregated in space to calculate the mean over the desired spatial unit for each pass through that spatial unit, typically referred to as a "single pass mean". The number of 1-s measurements for each spatial unit varies based on size (length or area), the speed limit of the street, and traffic conditions during the drive pass. This spatial aggregation allows multiple data points to be included in the calculation of the single pass mean, improving the estimate of the mean pollution level at that location. The use of the single pass mean for any particular spatial unit serves to give equal weight to each geographic portion of the drive regardless of how many 1-second data points were collected over that spatial unit. The result is time-resolved data at the geographic spatial unit level rather than 1-second.

The mean of the 1-second data points will be influenced by outliers, resulting in a collection of single-pass means that accommodate a greater degree of variability due to sampling than a collection of medians. Therefore, the resultant uncertainty estimates will be more conservative than if a single median had been selected.

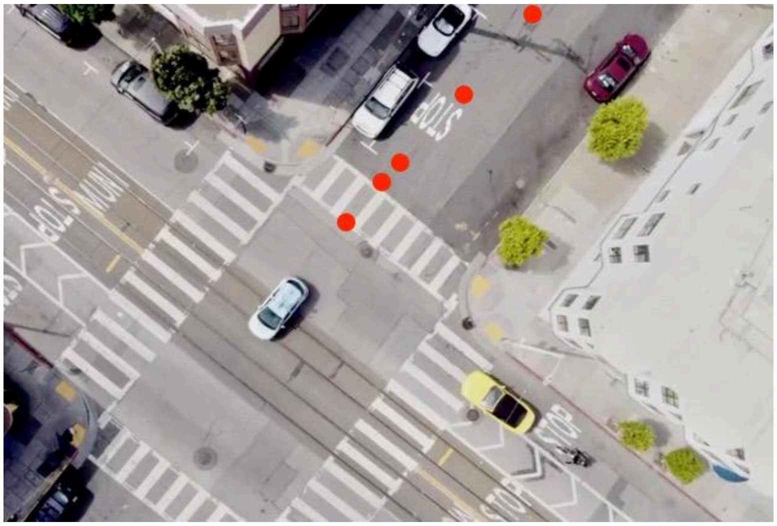

Figure 1. Illustration of 1-Hz data points as red dots aligned to the route of the car.

2.2 Data cleaning

Prior to using the 1-second data to produce ambient concentration estimates, data is flagged using specifically defined qualifier codes for conditions that, due to their sampling conditions or timing, are likely to result in biased estimates for typical ambient concentrations.

There may be times during routine driving when the car may sample its own tailpipe emissions, which we refer to as "self-pollution." Self-pollution will result in measured pollution concentrations higher than the local atmospheric concentrations for most pollutants (ozone would be biased lower) and is thus not representative of the "true" atmospheric conditions. Inclusion of data affected by vehicle self-pollution will bias the calculation of atmospheric concentration estimates. Note that we use hybrid cars in our fleets that typically shut off their gasoline engines when they stop and thus reduce possible instances of self-pollution. While electric vehicles are also used in our fleet, which are not subject to the same types of self pollution concerns, we use the same approach across all data collected.

We have identified two conditions where there is potential for self-pollution; (1) when a vehicle is stationary for an extended period of time or (2) repeated drives in a short period of time of the same section of road where the vehicle may pass through its own plume, such as might happen in a cul-de-sac or when making a u-turn. Whether measurable self-pollution does occur as a result of these two conditions, we conservatively remove segments that meet either of these two criteria from further analysis. We identify a vehicle as stationary when there are more than 75 seconds of data in a single road segment. The intent is not to remove typical idling situations found at stop lights/signs, but instead to remove a prolonged stay in a single location that may be due to a driver stopping for one reason or another. (typically we find that that just over 99% of all passes take less than 75 seconds.) In the second case, Aclima removes sequential passes of a segment within 30 seconds of the prior pass. Drivers have been trained to reduce the likelihood of self-pollution in these and other situations.

3.0 Ambient concentration estimates

Air pollution has significant variability over time and space, which means that the creation of geospatial, time-integrated maps from spatially-resolved measurements is complex. Both must be considered when designing analysis and data-driven modeling methods to produce ambient concentration estimates.

Pollution concentrations can vary over time scales that range from less than a minute to months. Short time scale variability, on the order of seconds to minutes, are most often observed near sources (e.g. the plume from a combustion source) but may also reflect spatial variability between neighborhoods or regions within a city. Within-day, day-to-day, and yearly variations in pollution concentrations can result from phenomena including changes in the temporal patterns of when emissions occur, atmospheric dynamics, seasonal changes in the weather, and changes in regional concentrations caused by synoptic-scale meteorology.

Spatial variability in air pollution over the scale of a city block, between neighborhoods, or between cities mainly arises from the location of and distance from sources, effects of urban design such as street canyons, and differences in microscale weather, like variations in temperature and wind speed and direction that arise from local and regional topology.

High spatial resolution average concentration estimates are derived from batch processing the individual measurements (either as 1 second measurements or single pass means) across the full geographic area interest and time window over which mapping occurred. The key challenge in the calculation of these high spatial resolution estimates is to separate the part of the time-resolved signal in a measurement area that comes from true spatial variability from that due to temporal changes that may be spatially homogeneous, but sampled at different times in different locations — sifting signals at multiple temporal and spatial scales in order to separate region-wide changes from hyperlocal signals of interest. The signal processing algorithms must also handle measurement correlations that are inherent in any set of observations.

We have designed an estimation framework to address separation of signals of interest as well as empirical spatial and temporal correlation in the road segment pollutant concentrations, tailoring algorithms to derive the best estimate of the average of each individual pollutant at high spatial resolution over the monitoring time period. These modeling strategies require a number of assumptions which require validation, which is described in Section 5.

3.1 Modeling strategies

Aclima's overall modeling approach for generating long-term average pollution estimates over the monitoring time period consists of the decomposition of the input observations into broad regional spatial trends, region-wide temporal trends and the hyperlocal signal. These components are then recombined into average estimates at hyperlocal spatial resolution.

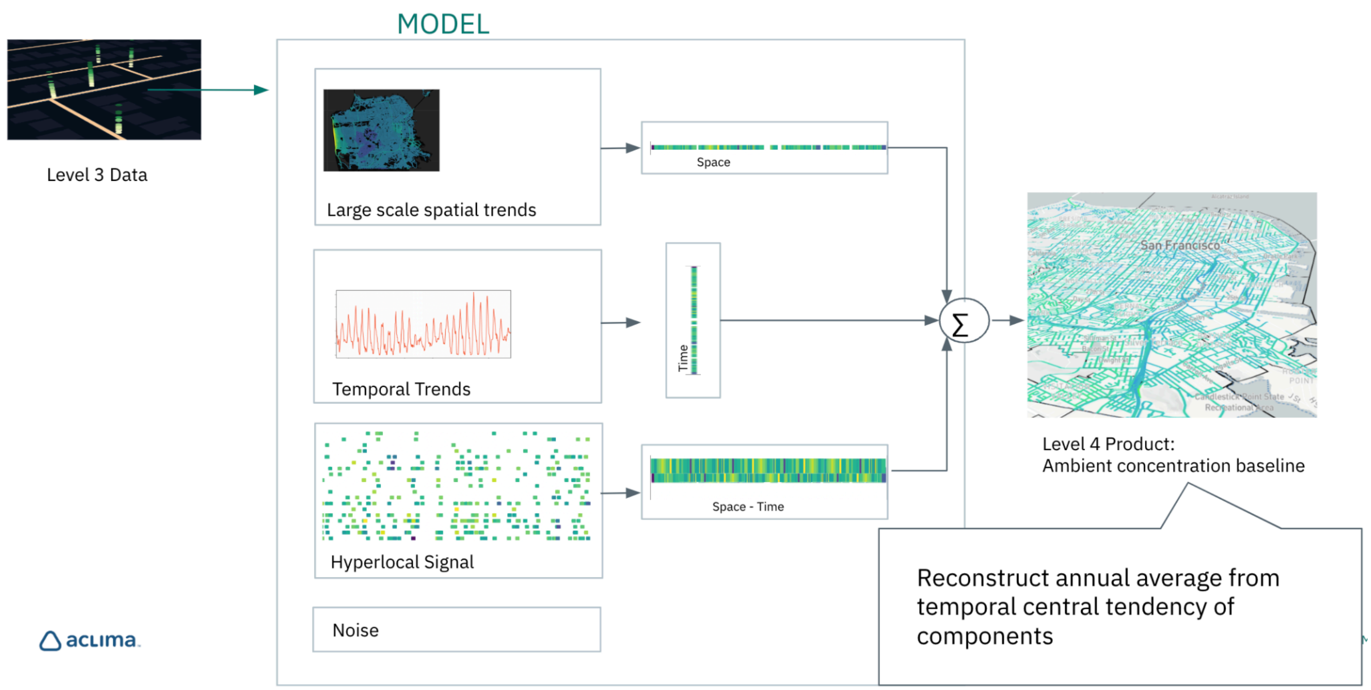

Figure 2 - Visual illustration of how ambient concentration estimates for each segment are produced by combining the central tendency of the large-scale spatial and temporal trends with the hyperlocal signal.

The constituent modules of this framework are tailored to each pollutant to ensure the best estimate of average for individual pollutant concentrations. The noise and structure of the input measurements dictate applicability of different estimation processes for decomposition and reconstruction. The data-driven models are introduced in the following sections.

3.1.1 Background normalized median method

Signal decomposition and reconstruction

For the background normalized median model, we bypass the large spatial trend decomposition and directly separate out the temporal trends across the region from the hyperlocal signal. A background model is formed from temporal trends by pooling information from across our mobile platform network.

For each single pass of the relevant spatial unit, the difference between the single pass mean concentration for that spatial unit and the hourly background measurement is calculated to form the hyperlocal signal. This difference reflects an enhancement or decrement relative to the background model during that hour. The observed positive or negative difference is added to the median pollutant level of the background model over the full monitoring time period, which can then be reconstructed with the median of the background signal, reconstructing the observations into a normalized signal with local differentials. The median in time across this reconstructed data for each segment yields the average estimate over the monitoring time period.

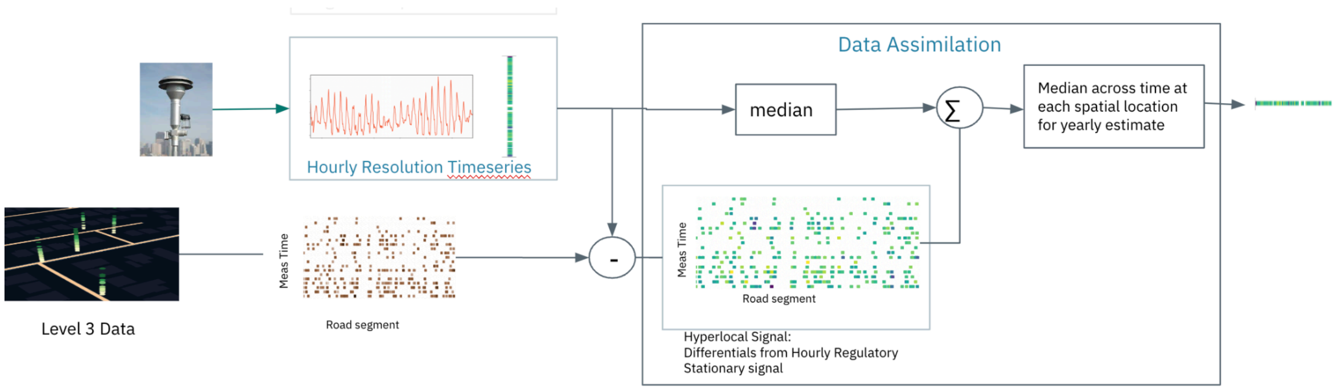

Figure 3: Visual illustration of the background normalized median process.

Background model

The methods used to generate background models may include identifying relevant regulatory monitoring stationary sites and deriving a background model directly from aggregated sensor measurements. For SMMI, because there is inconsistent availability of suitable regulatory monitors across all of the monitoring areas, sensor-based backgrounds will be used.

Sensor-based background signals are determined by pooling sensor measurements across time and extracting a percentile matched to pollutant behavior. Percentiles are chosen based on parameter tuning through comparison to regulatory sites (if available), noise of the sensors, and expectations of pollutant behavior. For example, when modeling pollutants in a small dense urban area, a high percentile for ozone concentrations represents background, as local emissions on top of the background result in dips in the ozone concentration.

Credible interval methodology

Estimating uncertainty for segment-level statistics for pollutant concentrations, specifically the Background Normalized Median, does not lend itself well to the most common statistical procedures for computing confidence intervals. Low sample counts, non-gaussian concentration distributions, and measurement noise makes those methods inappropriate and difficult to interpret. Aclima uses the Bayesian Bootstrap approach to estimate the uncertainty associated with each spatial unit over the monitoring period. This method randomly resamples the initial dataset, resulting in a probability distribution over plausible values of the median given the observed data. This probability distribution is termed the posterior distribution in Bayesian statistics and allows us to directly estimate uncertainty intervals for estimates of typical pollutant levels during the monitoring period. This estimate of uncertainty or credible interval is the Bayesian statistics cousin to the confidence interval.

The Bayesian Bootstrap procedure is performed to generate posterior distributions for the spatial unit BN Median, from which uncertainty estimates can be calculated. The process utilizes a uniform Dirichlet distribution, which models the randomness of the probability of outcomes parameterized by a vector of positive valued numbers based on the underlying single pass data for an individual spatial unit and a number of passes. A single Bayesian Bootstrap replicate uses the Dirichlet distribution to sample the probability weights (alpha) for the pass data for that spatial unit. To compute posterior distributions for the BNMedian, we use the weighted median from a number of draws. Once posterior distributions for the BNMedian (weighted median) are obtained, 95 percent credible intervals are calculated and reported as uncertainty estimates.

It is important to note that the confidence intervals are designed to capture the precision of the estimate but are not appropriate tools for quantifying systematic bias that could influence the accuracy of the pollution estimates. Examples of systematic bias include device-level bias for individual sensors and potential inaccuracies in the sensor models. At the device level, systematic bias is better captured by pre- and post-deployment calibrations and the associated acceptance criteria.

3.1.2 Statistical measurement reconstruction method

Signal decomposition and reconstruction

A statistical method is used to produce ambient concentration estimates based on correlations in the spatial and temporal measurements obtained during mapping. The method is designed to take a data set that is sparse in space and time and generate estimates of likely pollution levels in all locations and at all times, filling in the gaps. The method imputes pollution concentration data for all locations for all days, which are averaged over the measurement period to obtain a single estimate of the ambient concentration for each location at the desired spatial unit.

A general additive model is used to identify the large-scale spatial and temporal trends in the collected mobile monitoring data set. This general additive model relies on establishing correlations between the mobile monitoring pollutant data with external data sources, including information about road type, topographical data, time of day, time elapsed since the start of monitoring, and gridded meteorological reconstruction data (NOAA's High Resolution Rapid Refresh model). Once these correlations are established, they are used to interpolate the spatially and temporally sparse mobile monitoring data set across the entire space and time domain for the monitoring area and time period. This allows for a consistent representation of the large-scale atmospheric trends in space and time to be consistently applied across the monitoring domain in order to minimize sampling bias in the final ambient concentration estimates. Examples of these large scale features include, for example, seasonal or diurnal trends in air pollutants, large scale pollution events, such as long range transport of wildfire smoke, differences in pollutant concentrations with elevation, and differences in concentrations with road types (highways vs residential roads, for example)

The data residuals remaining after decomposing the collected data into these correlated trends is considered the "hyperlocal" signal, which is still temporally and spatially sparse. This residual signal is influenced by local scale air pollutant variability, for example plumes of pollution from individual sources that are only detected in very localized areas, such as individual city blocks. This data is spatially and temporally interpolated using two commonly used geospatial analysis methods: DINEOF (Data INterpolation with Empirical Orthogonal Functions; Alvera-Azcárate et al. 2011) in combination with a Kalman Filter (Akatsuka, 2023). A limitation of this method is that highly localized discrete air pollution events (i.e. facility breakdowns, local fires, etc.) may not be captured well by this procedure. The advantage, however, is that the spatial distribution of higher and lower concentrations shown in the resulting map of ambient concentrations are less likely to be influenced by bias due to temporal sampling bias.

Credible interval calculation
The result of the combined methods is a distribution of estimates for each segment and each day. The ambient concentration estimate reported is taken as the central tendency (mean or median) from a distribution of values generated through multiple trials from the set of daily data. The credible interval is calculated from the spread of this distribution.

3.1.3 Pollutant Method Mapping

The chart below lists the methods employed for each pollutant from the Aclima Mobile Platform. Multiple methods for a single pollutant indicates an average concentration estimate product has been produced using each of these methods since 2019. Aclima determines the best-suited model for producing the ambient concentration estimate depending on the pollutant and availability of relevant data. The method for validation of these approaches is discussed in Section 5.

Pollutant	Background Normalized Median	Statistical Measurement Reconstruction	Median
CO	X	X	
CO_2	X	X	
$PM_{2.5}$		X	
O_3	X	X	
NO_2	X	X	
Black Carbon	X		X
Various Air Toxics[2]	X		X

4.0 Metrics for collection adequacy

The degree to which our ambient concentration estimates are representative of typical concentrations observed over the sampling period depends on (1) having sufficient observations (i.e., average repeat visits to roads in the mapping area) over the contract time period and (2) that those observations are sufficiently distributed across the measurement period to account for the intrinsic variation of pollutant levels.

Alicma uses a dynamic sampling algorithm that is updated daily with the goal of collecting data that maximizes improvement in the characterization of air quality rather than specify a number of samples on any individual length of road. The system ensures sufficient data collection to support spatially resolved ambient concentration estimates, with sampling deliberately distributed to provide higher rates of repeat measurements in locations with higher observed variability. The driving algorithm is designed to complete an average of 20 repeat measurements distributed across all residential and major roads in all census block groups. Twenty repeat measurements is generally found to be the point at which additional repeats only provide marginal reductions in uncertainty(Apte et. al. (2017). The dynamic sampling algorithm accounts for different locations requiring different numbers of repeat measurements to achieve this in order to optimize the use of monitoring resources. More detail on the algorithm can be found in the Aclima Mobile Measurement Quality Assurance System, Section 3.1).

[2] Air Toxics measured by the SMMI Partner Mobile laboratories may be used to generate ambient concentration estimates.

Drive passes identified as influenced by self-pollution (Section 2.2) are not included as part of the pass count.

5.0 Validation of segment aggregate concentration estimates

The unique value of hyperlocal maps resulting from mobile mapping also makes validation difficult as independent data at this spatial resolution are not available for comparison. To address this, we have developed a variety of approaches to validate the performance of our ambient concentration estimates. This section presents the results of these approaches using historical data aggregated to a 100-m road segment spatial aggregate.

5.1 Validation by comparison with regulatory measurements

Quantifying uncertainty at the device level is a key piece in understanding data quality of the hyperlocal maps, but the uncertainties do not necessarily propagate in a straightforward way to the final data products. For this reason, Aclima's approach is to compare our mobile measurements to stationary measurements (e.g., regulatory reference) where those data are available in order to directly quantify uncertainty against the traditional, established methods for measuring ambient concentrations of pollutants over time. In regions where a suitable number of reference sites are available, the mobile-to-stationary comparisons can provide basic statistics to describe uncertainty broadly across the mapping region, extrapolating to locations where reference sites do not exist.

We have taken two approaches to map evaluation based on regulatory sites. These include, (1) a time-resolved approach where individual segment pass means are compared with the appropriate hourly averaged data (Section 5.1.1) and, (2) a time-integrated approach where ambient concentration estimates are compared with the median value reported by the station over the same time period (Section 5.1.2). The first approach is an extension of the device-level data quality evaluation, but allows for an aggregation of bias across all devices contributing to the hyperlocal map. Additionally, it allows for a determination of inter-network differences that could result from, for example, systematic differences in the calibration sources used or differences between different measurement techniques (i.e. optical particle detection vs gravimetric detection of $PM_{2.5}$). The second approach provides a quantification of overall uncertainty of the final typical concentrations, including device-level uncertainty, sampling uncertainty, and uncertainty resulting from the modeling approach used to produce the ambient concentration estimates. The results reported here are based on mobile mapping conducted across California between 2019 and 2021 and use the regulatory data reported to California Air Quality Management District's Air Quality Monitoring Information System (CARB AQMIS) over the same time period. As Aclima continues to map in more regions, we expect to refine these uncertainty estimates and determine how relevant these results are for other locations, including areas with a limited existing air quality measurement network.

Regulatory site measurements can be spatially representative over scales of several meters to several kilometers, depending on the site type as defined by USEPA (page 5). However, direct comparisons between mobile measurements and regulatory site measurements can be complicated due to the fact that mobile measurements represent on-road conditions, having high variability across different road types. As a result, it is important to consider the spatial and temporal aggregation scales as well as the maximum allowable distance between stationary and mobile concentrations, which we will refer to as the *distance buffer* in the following discussion.

Previous in-depth analysis (Whitehill et al., 2024, LaFranchi et al., 2022, Solomon et al., 2020) has shown that the key to making a meaningful comparison between mobile and stationary measurements is to reduce random noise in the comparison due to atmospheric variability by maximizing the number of collocation samples while minimizing the distance buffer and filtering out high variability road types (e.g. highways). Given sufficient collocation samples within a small radius, it is possible to accurately characterize bias between on-road measurements (both time-resolved and time-integrated) and a stationary reference network. While we have found (Whitehill et al., 2024) that there can be high degrees of correlation between mobile and stationary measurements even at distance buffers of 3-5 km, we choose to use a radius of 250m for this evaluation to reduce the likelihood of spatial variability influencing the results.

5.1.1 Time resolved comparison of mobile sensor measurements to regulatory measurements

To evaluate the quality of AMN measurements on a day to day basis, Aclima compared single pass mean segment data for $PM_{2.5}$, NO_2, O_3, and CO collected within a 250 m radius around a regulatory site to the data reported by that site. The single pass mean concentrations measured within this radius for each pollutant were averaged to 24 hours across all devices and sites and compared to the daily mean of the measurements from the regulatory site for the hours when the car was within the distance buffer. For instance, if the car was near the site in the 10 AM, 2 PM, and 11 PM hours, only the corresponding data from the same hours was used from the regulatory site. Note that more than one vehicle/device may be part of the comparison for any given day, and the overall temporal comparison will certainly be made up of data from multiple cars. This comparison provides an indication of the general quality of the AMN device data *in situ* and across many devices, as well as insight about the value of data aggregated over different spatial and temporal scales.

For this evaluation, we focus on data collected in the Bay Area Air Quality Management District (BAAQMD) region. The data from all cars and relevant sites in the region (i.e. the sites reported measurements for the pollutant and had roads within 250 m of the site) were averaged together for a network-wide assessment. The number of regulatory sites used in the comparison ranged from 14 to 17 depending on the pollutant. Figure 5 presents time series and regression analysis for $PM_{2.5}$, NO_2, O_3, and CO for the daily average values. Excellent daily average temporal agreement ($R^2 > 0.9$) with little bias (<8%) is observed for $PM_{2.5}$ and O_3. Agreement for NO_2 and CO is slightly reduced ($R^2 > 0.65$) with a bias of less than 20% for both. These results are comparable to other more direct estimates of device-level data quality from our pre- and post-deployment calibration audits as well as from an independent side-by-side collocation study at the Laney College monitoring site in Oakland, CA.

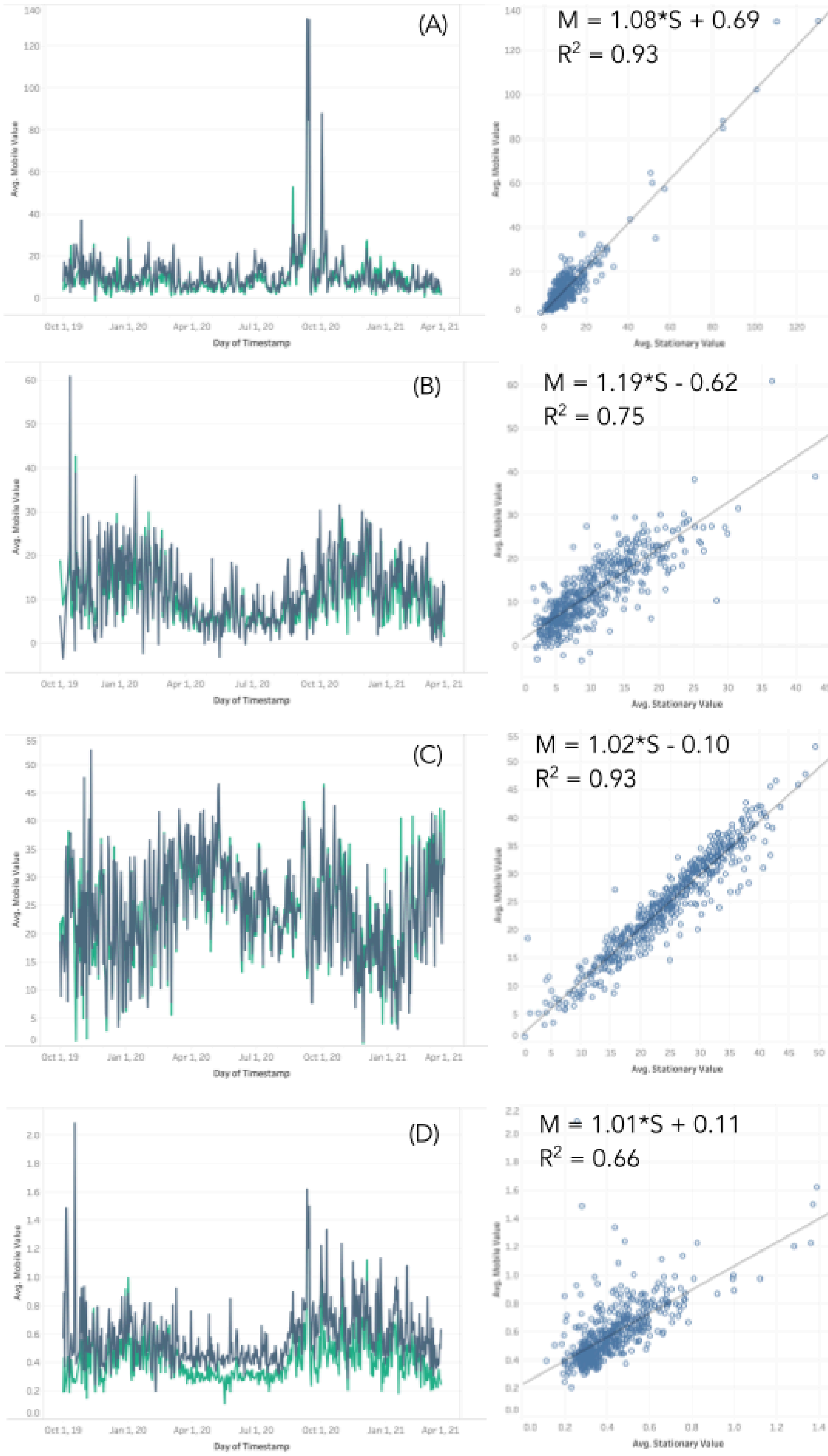

Figure 5. Comparison of daily mean mobile sensor (M) measurements within a 250 m radius circle centered at a stationary regulatory site to daily mean hourly data at that regulatory site (S). Data for all mobile platform measurements within 250 m of a regulatory site are compared to that regulatory site within the baseline period. Daily time series (left column) and recession analysis with regression statistics

(right column) are presented for A) $PM_{2.5}$ [μg/m^3]; B NO_2 [ppb]; C) O_3 [ppb]; and D) CO [ppm]. A Type II - Major Axis regression was used with no weighting.

These results, in relation to FRM and FEM measurements of precisely known uncertainty at stationary regulatory sites, confirms that Aclima's mobile platforms achieve high quality data at hyperlocal, block-by-block and community wide spatial scales with acceptable bias and very good temporal agreement. For example, based on an extensive literature search and international sensor workshops, the USEPA has developed a draft set of target values for air pollution sensors that measure $PM_{2.5}$ and O_3 (Williams et al., 2019; Duvall et al., 2020). Performance metrics reported (Duvall et al., 2021) included precision, bias and intercept, coefficient of determination (R^2), and error (RMSE) for $PM_{2.5}$ and O_3. Aclima's mobile sensor data quality objectives meet or exceed EPA's performance guidance for $PM_{2.5}$ and O_3. USEPA has evaluated sensor performance for NO_2, CO (Duvall et al. 2021) but has not formally summarized findings as done with $PM_{2.5}$ and O_3.

5.1.2 Time integrated comparison of ambient concentration estimates to regulatory measurements

To evaluate the overall performance of the Ambient Concentration Data Product, the ambient concentration estimates (ACE) at 100 m road segment aggregations within a 250 m radius of a regulatory site were compared to the annual median concentration from that regulatory site for each pollutant over the same date range of mobile monitoring collection (e.g. for an annual data collection, aggregating all data collected from June through May). (Note: BC is not included in this example analysis as there were not sufficient mobile BC measurements within 250 m of a regulatory station that also report BC.) Here we present an example of this evaluation using data collected throughout California, including the BAAQMD region as well as neighborhoods in San Diego, Sacramento, Los Angeles, and San Bernardino. We perform this same analysis whenever we complete mapping.

Table 2 presents several evaluation metrics for the combination of all regulatory sites and all of the data collected throughout California based on the observed differences ($D_{j,s}$) between each segment (s) and stationary site (j). This collection of $D_{j,s}$ values are aggregated to calculate a mean ($\overline{D}_j$) and a standard deviation ($\sigma_{D,j}$) for each site. The evaluation metrics based on $\overline{D}_j$ are then defined as follows:

- ***Mean Bias Error (MBE)***, which provides an estimate of systematic bias between our ACE and regulatory measurements

$$MBE = \frac{\sum_{j=1}^{N} (\overline{D}_j)}{N}$$

- ***Mean Absolute Error (MAE)***, which provides an estimate of absolute bias between our ACE and regulatory measurements across all sites

$$MAE = \frac{\sum_{j=1}^{N} (|\overline{D}_j|)}{N}$$

- ***Centered Root Mean Square Error (CRMSE)***, which is a bias-adjusted version of Root Mean Square Error, and provides an estimate of precision of the ACE values compared to regulatory measurements, where outlier differences are weighted more heavily.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N} (\overline{D}_j)^2}{N}}$$

$$CRMSE = \sqrt{RMSE^2 - MBE^2}$$

- ***Correlation Coefficient (R^2)***, determined from a linear regression (OLS) between the mean ACE concentration around each site ($\overline{ACE}_j$) and the stationary median values for each site, j, which provides a measure of the ability of the ACE map to reproduce the variance observed across the stationary network.
- ***Standard Deviation (σ_D)***, which describes the variability at the segment level around each site, reflecting true local variability due to different road types and local sources as well as random precision uncertainty for ACE values at segment aggregations

$$\sigma_D = \frac{\sum_{j=1}^{N} (\sigma_{D,j})}{N}$$

Table 2: Performance metrics for comparison of map segment aggregates within 250 m of a regulatory site to that site for all regulatory site locations and baselines periods.

	MBE	MAE	CRMSE	R^2	σ_D
$PM_{2.5}$	+0.5 µg/m3 (11%)	1.7 µg/m3 (23%)	2.2 µg/m3 (30%)	0.40	0.8 µg/m3
NO_2	-0.2 ppb (-5%)	1.7 ppb (30%/22%*)	2.2 ppb (44%/28%*)	0.77	2.0 ppb
O_3	+0.4 ppb (+2%)	1.2 ppb (5%)	1.4 ppb (6%)	0.82	0.8 ppb
CO	+0.04 ppm (+15%)	0.06 ppm (21%)	0.07 ppm (25%)	0.11	0.04 ppm

* NO_2 % MAE and % CRMSE values are heavily influenced by several sites where NO_2 concentrations are relatively low. When excluding sites with annual median concentration less than twice the CRMSE value, the % MAE and % CRMSE reduce to 22% and 28%, respectively.

The results in Table 2 can be interpreted as overall uncertainty for the ambient concentration estimates combining uncertainties across the entire platform, including device level, sampling, and modeling uncertainties. In order to help put these results in perspective, these metrics are also calculated as a percentage using relative differences for each site and compared to the uncertainty thresholds listed for different use cases according to the EPA Air Sensor Guidebook (Williams et al, 2014), as shown in Table 3. Bias (as MAE) ranges from 5% (O_3) to 30% (NO_2). Systematic bias (as MBE) is relatively low, ranging from -5% (NO_2) to +15% (CO), indicating that most sources of uncertainty in the ACE data product are random.

Precision (as CRMSE) ranges from 6% (O_3) to 44% (NO_2). As a percent, NO_2 bias and precision are somewhat high (~30-45%), however, this is driven by high relative uncertainties at several sites where annual median concentrations are less than 5 ppb. Excluding these low NO_2 sites, the % precision and bias are both <30%. The Hotspot Identification (Tier II) use case for all 4 criteria pollutants is achievable. For O_3, the % bias and % precision are low enough to be used for the Supplemental Monitoring use case (<20%).

Table 3: Adapted from the EPA Air Sensor Guidebook (Williams et al., 2014), showing the different tiered use cases for sensors according to precision and bias uncertainty alongside the ACE Data Product modalities suitable for each tier. Tiers where the ACE data products do not support the use case are listed as not applicable (NA).

Tier	Use Case	Precision and Bias Uncertainty	ACE Data Product Modalities
I	Education and Information	<50%	NO_2, O_3, $PM_{2.5}$, CO
II	Hotspot Identification and Characterization	<30%	NO_2*, O_3, $PM_{2.5}$, CO
III	Supplemental Monitoring	<20%	O_3
IV	Personal Exposure	<30%	NA
V	Regulatory	O_3 (<7%) $PM_{2.5}$ (<10%) NO_2 (<15%) CO (<10%)	NA

* In locations where NO_2 ACEs are higher than ~5 ppb.

In order to further illustrate the comparisons between ACE concentrations and the stationary site concentrations, Figure 6 shows the scatter plot between ACE and stationary sites, with error bars in the figure representing $\sigma_{D,j}$ for the collection of segments around each $\overline{ACE}_j$ value. The best fit lines shown are ordinary least squares fit to the data. as a function of the median regulatory site concentration. The correlations from these plots between the data are shown in Table 5 as R^2 calculated using ordinary least squares. In addition, the standard deviation for the set of segments around each location are also displayed around each segment mean value as a measure of segment to segment variability around each site.

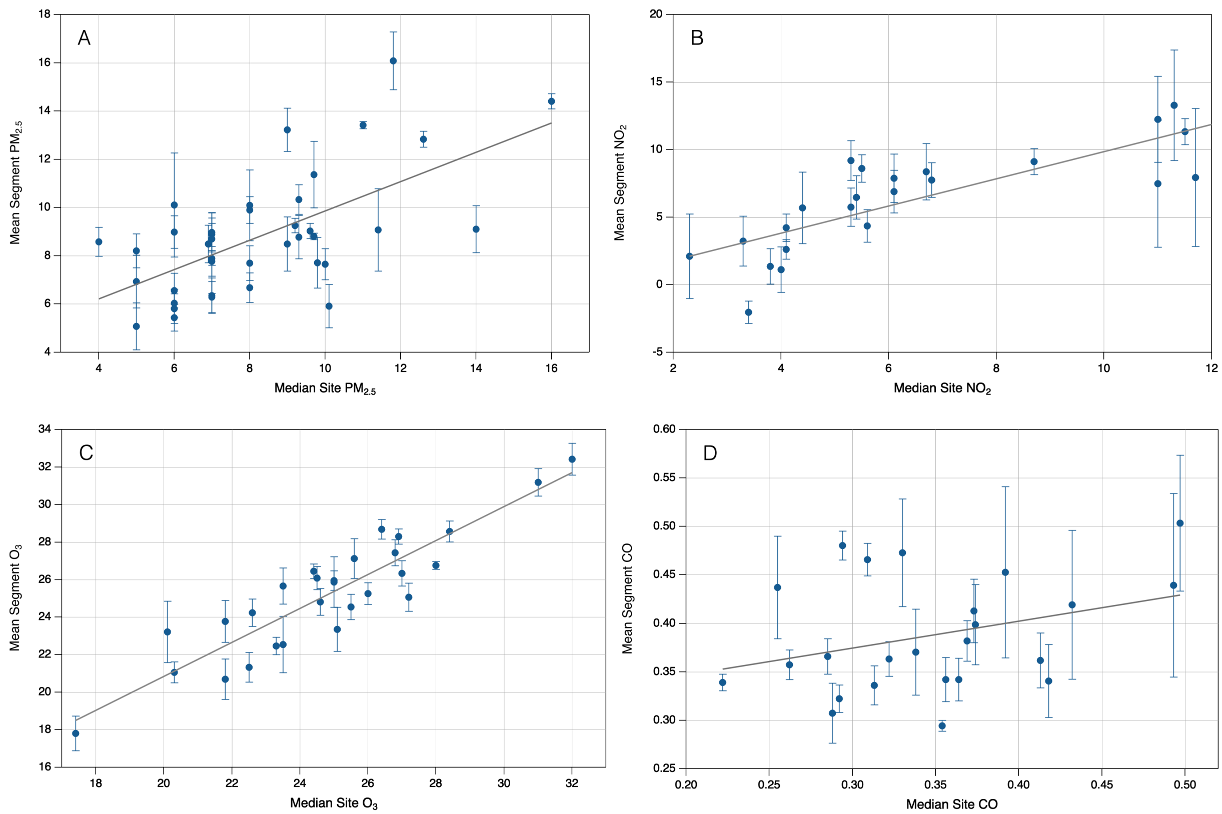

Figure 6. Comparison of the baseline annual mean segment concentration within 250 m from the regulatory site annual median for each site for which a comparison was possible for (A) $PM_{2.5}$ (μg/m^3), (B) NO_2 (ppb), (C) O_3 (ppb), and (D) CO (ppm). The standard deviation of the set of segments within the 250 m radius that comprise the mean are shown as error bars around the mean value. The trend line is the result from an OLS linear fit to the data.

The figures show general agreement with the trend line, with O_3 and NO_2 showing strong correlations, with R2 of approximately 0.8 for both, while CO has poor correlation (R2 = 0.1) and $PM_{2.5}$ has moderate correlation (R2 = 0.4). The precision uncertainty (as CRMSE) for CO is approximately the same order of magnitude (0.07 ppm) as the standard deviation of CO concentrations observed across all of the regulatory sites used in this analysis. Similarly, $PM_{2.5}$

variability across the sites used in the analysis is relatively small (2.6 ug/m3) compared to the CRMSE of 2.2 ug/m3. As a result, the R2 values for both CO and $PM_{2.5}$ likely do not reflect the true performance that would be found in comparison with stationary networks capturing a wider range of concentrations. As Aclima continues to collect data in more diverse locations, we will continue to refine these uncertainty estimates and improve confidence that they are generalizable to more locations.

This analysis is not currently possible for CH_4, CO_2, C_2H_6, and BC because of data availability for these pollutants at suitable stationary sites in locations that Aclima has mapped to date.

5.2 Accounting for Systematic Measurement Bias

In the examples provided in Section 5.1, the bias between the Ambient Concentration Estimates and existing regulatory measurements was relatively low for all pollutants evaluated. While a robust and well-executed measurement quality assurance system is key to minimizing this systematic bias, the possibility for significant bias is always a possibility when comparing two different measurement networks operated by different organizations using different primary standards, detection methods, and levels of data quality. We continue to assess Aclima sensor performance compared to different reference equipment, with an ongoing program of research and development to understand how accuracy varies in different locations with varying pollution sources and concentrations, meteorology, and other factors. Sensors in the platform not routinely compared directly to a reference method (CO, PM2.5, and Black Carbon) as part of its standard calibration procedure may be particularly prone to systematic bias.

In addition to the approach described in Section 5.1.1 of using time-resolved comparisons between mobile and regulatory measurements to evaluate the quality of AMN measurements, direct collocations of Aclima's AMN and associated sensors at existing monitoring sites is an additional approach that can be used to increase confidence in the characterization of sensor bias. In addition, this approach can help identify any time-dependent correlations with the degree of bias that might be missed from the mobile vs stationary comparisons (i.e. seasonal, time of day, with varying meteorological conditions etc.) as well as provide valuable context for interpreting bias resulting from the fleet-wide comparisons of mobile measurements to stationary measurements.

At the end of a measurement period for the ambient concentration estimates, Aclima has the option to use these in situ comparisons to adjust for systematic bias in certain cases. This process helps to better harmonize Aclima's measurements with existing measurement networks, which are usually the best source of truth for a given pollutant. While this approach may be applied to any pollutant, a typical scenario where this is expected to be necessary is for PM2.5. The parameters used in the sensor model that converts Aclima's particle count measurements to PM2.5 have been found to vary across different geographies, attributed to differences in size distribution and chemical composition of the ambient aerosol in these locations. Additionally, there are known sources of bias even between different approved (Federal Equivalent Methods or FEM) regulatory methods for measuring PM2.5, and Aclima's PM2.5 sensor has been found to have different degrees of systematic bias when compared to different FEM methods of measuring PM2.5. In cases where these sources of systematic bias are deemed to be significant, the in situ comparisons are used to derive the optimal set of parameters to apply for a particular geographic location prior to generating the final ambient concentration estimates.

5.3 Additional validation of model-generated ambient concentration estimates

Aclima's modeling strategies are validated through a number of approaches. In addition to the comparisons with stationary monitors described above, we employ a series of test scenes to assess the ability of a model to capture specific types of intrinsic spatial and temporal features in air pollution concentrations, such as sharp spatial gradients in concentrations. We additionally use self-contained validation strategies for calculating errors with subsets of the measurements against model predictions, as well as for assessing model stability.

5.3.1 Evaluation using test scenes

We test our models against a series of test scenes, much like vicarious calibration of remote sensing systems against targets with known, constant surface reflectance. We consider the model as encompassing both the sampling strategy (the true trajectories and measurement times from our fleet in the region) as well as the algorithms for estimating concentrations since these are naturally coupled. These test scenes are not meant to capture the full complexity of the atmosphere and resulting patterns in pollutant concentrations. Instead, they are intentionally selected to stress test a model against specific observable features in a repeatable manner. These test scenes enable calculation of the effective resolution of the model against a broad set of signals.

5.3.2 Self-contained validation strategies

Self-contained validation strategies use subsets of the measurement against model prediction to calculate errors, and can also be used for assessing model stability. Our self-contained validation strategies include internal goodness of fit, parameter sensitivity training, out of sample validation, and validation against third party data.

6.0 References

Alvera-Azcárate, A.; Barth, A.; Sirjacobs, D.; Lenartz, F.; Beckers, J.-M. Data Interpolating Empirical Orthogonal Functions (DINEOF): A Tool for Geophysical Data Analyses. Medit. Mar. Sci. Spec. Issue 2011, 5–11.

Apte, J.S., Messier, K.P., Gani, S., Brauer, M., Kirchstetter, T.W., Lunden, M.M., Marshall, J.D., Portier, C.J., Vermeulen, R.C.H., and Hamburg, S. 2017. High-resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. Environ. Sci. Technol. 51, 6999–7008.

Duvall, R., Clements, A., and Goodman, L. 2020. Overview of the U.S. EPA Work on Developing Performance Testing Protocols and Targets for Ozone and PM 2.5 Air Sensors. Presented at the Air Sensors International Conference, Virtual Series, October 29, 2020.

Duvall, R. M., Hagler, G. S. W., Clements, A. L., Benedict, K., Barkjohn, K., Kilaru, V., Hanley, T., Watkins, N., 1045 Kaufman, A., Kamal, A., Reece, S., Fransioli, P., Gerboles, M., Gillerman, G., Habre, R., Hannigan, M., Ning, Z., Papapostolou, V., Pope, R., Quintana, P. J. E., and Lam Snyder, J. 2021. Deliberating Performance Targets: Follow-on workshop discussing PM_{10}, NO_2, CO, and SO2 air sensor targets, Atmospheric Environment, 118099.

Kock, J., Demirel, M.C., and Stisen, S., 2018. The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models. Geosci. Model Dev., 11, 1873-1886,

LaFranchi, B., Ridley, D., Trickett, C., Kwasnica, M., Chow, M., Lunden, M., George, K., Smith, J, and Vona, M. Calibration of citizen sensor networks using a mobile air monitoring platform. Presented at the 2022 Air Sensors International Conference, May 11, 2022, Pasadena, CA.

Solomon, P. A., Vallano, D., Lunden, M., LaFranchi, B., Blanchard, C. L., and Shaw, S. L. 2020. Mobile-platform measurement of air pollutant concentrations in California: Performance Assessment, Statistical Methods for Evaluating Spatial Variations, and Spatial Representativeness. Atmos. Meas. Tech. 13, 3277–3301.

Williams, R., Vasu Kilaru, E. Snyder, A. Kaufman, T. Dye, A. Rutter, A. Russell, AND H. Hafner. Air Sensor Guidebook. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-14/159 (NTIS PB2015-100610), 2014.

Williams, R., R. Duvall, V. Kilaru, G. Hagler, L. Hassinger, K. Benedict, J. Rice, A. Kaufman, R. Judge, G. Pierce, et al, 2019. Deliberating performance targets workshop: Potential paths for emerging $PM_{2.5}$ and O_3 air sensor progress. Atmos. Environ. X 2:100031.

Whitehill, A.R., Lunden, M.M., LaFranch, B., Kaushik, S., and Solomon, P.A. 2024 Mobile air quality monitoring and comparability to stationary monitoring sites for quality assurance during mobile deployments. Atmos. Meas. Tech. 17, 2991-3009..

Akatsuka, H. and Terada, M. Application of Kalman Filter to Large-scale Geospatial Data: Modeling Population Dynamics. ACM Trans. Spatial Algorithms Syst. 9, 1, Article 6 (March 2023).