Community Air Monitoring Plan: Appendix E

California Statewide Mobile Monitoring Initiative (SMMI) Hyperlocal Enhancement-Based Data Products Quality Assurance System (v2.2)

aclima

August 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 4
2.0 Background and Methodology 6
2.1 Enhancement Event Detection 6
2.2 Enhancement Event Methodology 8
2.2.1 Deriving the Baseline 8
2.3.2 Enhancement Event Identification 9
2.2.3 Enhancement Event Processing 10
2.2.4 Summary of Critical Input Parameters 10
2.3 Enhancement Clusters 11
3.0 Impact of Sensor Uncertainty 12
4.0 General Use Cases and Limitations 15
4.1 Use Cases 15
4.2 Limitations 15
5.0 Citations 16
7.0 Appendix 17
7.1 Definitions 17

1.0 Introduction

Mobile air pollution mapping is a flexible method to measure concentrations of a broad range of air pollutants and greenhouse gases over large geographic areas at high spatial resolution. A foundational data product Aclima produces from mobile mapping is hyperlocal maps of typical pollution concentration over a defined time period that illustrate areas of higher or lower concentration at the street level. This ambient concentration data product is intended to produce long-term average ambient concentration estimates that would be comparable to an average stationary measurement made nearby. As such, the ambient concentration data product can minimize the influence of large, short-lived increases in concentration that can be observed when measuring near emission sources. The detection of these higher-than-average measurements in the time-resolved data is an indication of a localized emission source, and analysis of these signals can more directly facilitate the identification and characterization of sources at hyperlocal spatial scales.

As the mobile platform moves through an emissions plume, a temporary increase in concentration may be observed over time, which we call an enhancement event. We define these enhancement events as localized elevation in concentration of a given pollutant within the plume that is measurably distinguishable from the ambient background (as measured in a nearby location at close to the same time). These individual enhancement events represent the detection of an emission event for a particular pollutant in the immediate vicinity of the source. Collectively, these enhancement events are compiled and spatially aggregated into localized clusters that can be used to identify the locations of air pollution sources.

These types of analyses are designed to meet the following monitoring objectives:

- Characterize the hyperlocal spatial distribution of pollutant emissions sources
- Provide information on the type of source leading to pollutant enhancements
- Assess the persistence of that pollutant source in any given location
- Provide a qualitative to semi-quantitative indication of the magnitude of enhancements observed in any given location, within the limitations of the sensor measurement

Two features of the SMMI broad area monitoring approach are particularly well-suited to support these enhancement-based objectives: 1) simultaneous measurement of multiple pollutants and 2) the multipass sampling approach.

Targeted area monitoring is also well suited to support the above objectives. It is important to note that, because targeted area monitoring takes place over shorter time periods (i.e. on the order of two weeks), the results are not sufficient to characterize the long-term persistence of pollutants. However, the SMMI Partner Mobile Labs have the ability to do a more detailed chemical characterization of enhancement events, which strengthens source identification capability and supports improved understanding of the potential risk associated with individual pollutant sources. (For more detail on SMMI Partner Mobile Labs QA procedures, see CAMP Appendix G.)

Simultaneous measurement of multiple pollutants allows for additional specificity in identifying the source of the emission that resulted in the enhancement. For example, measuring methane (CH_4) and ethane (C_2H_6) simultaneously provides the means to distinguish between biogenic and thermogenic (i.e. fossil) sources of methane. Multipass sampling adds a critical dimension for analysis: persistence. Persistence is defined as the consistency over time with which a specific location is impacted by a particular emissions source. High persistence of enhancement events can be important for increasing confidence in the presence of a particular source in a specific location as it means that the enhancement cluster is unlikely to be a result of a sporadic spike due to sensor noise (it is highly unlikely that a sporadic spike would happen in the same location on multiple occasions) and is unlikely to be due to the detection of a highly transient source (i.e. a single highly-polluting vehicle passing near the mobile platform in an otherwise unpolluted location).

The Data Quality Objectives for the use of enhancement events to characterize air pollutant emission sources are as follows:

- Find and map spots where pollution is likely coming from by detecting noticeable spikes in measurement readings that are clearly above normal background levels. More specifically, this means that the spike measurement must have a signal to noise ratio of at least 3, where “noise” is defined as the 1 second precision of the measurement.
- Ensure high confidence in the locations where pollution emissions sources are detected by minimizing the presence of “false positives.” False positives are minimized by making sure that multiple detections of emissions sources occur in the same location before identifying it as a likely source of pollution and is typically quantified as the number of detections per visit to a particular location.
- Monitor and track the performance of each underlying measurement using the key data quality indicators of gain, drift, and limit of detection

The Aclima Mobile Platform and Partner Mobile Laboratories support the source identification for the following pollutants:

- CH_4, C_2H_6, BC, $PM_{2.5}$, NO, CO, TVOC, toxic air contaminants

In the following sections, we provide a detailed explanation of the data processing methodology, descriptions of the quality assurance processes behind data product design, a description of the relevant quality assurance processes behind data collection (which are discussed in more detail elsewhere), a discussion of data products limitations and use cases, and details on each individual data product along with validation examples for some of their intended use cases. This methodology is generalizable and can apply to both broad area monitoring and targeted area monitoring, Aclima mobile platform data and Partner Mobile Laboratory data.

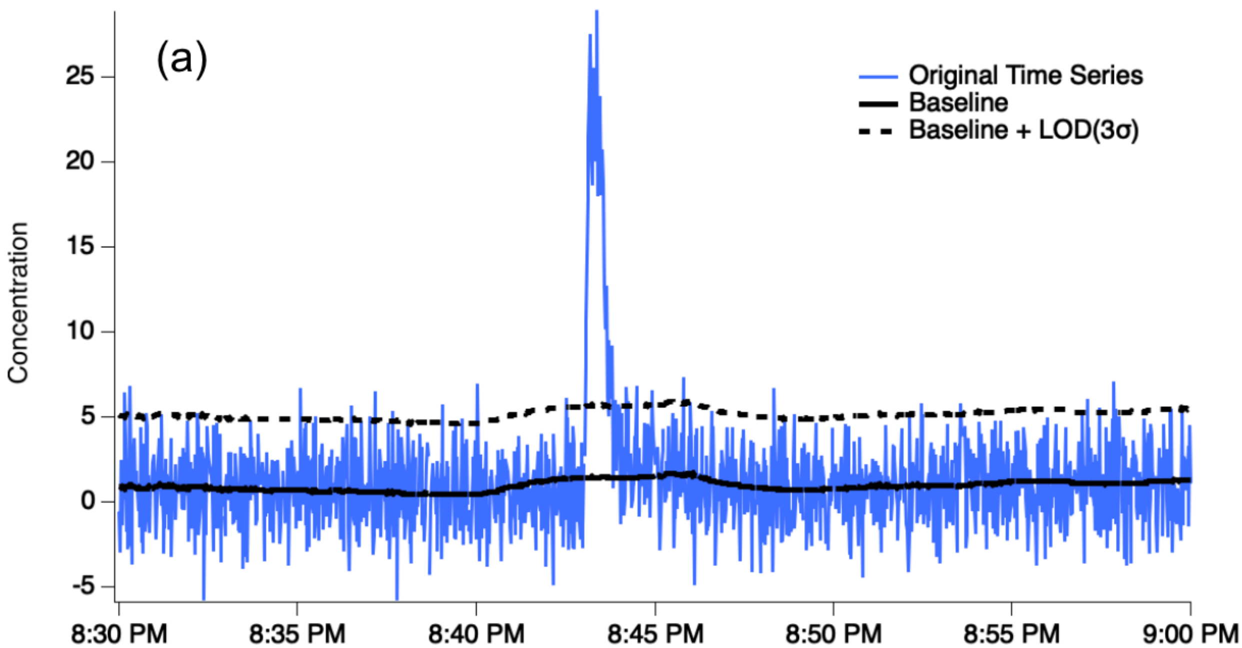
2.0 Background and Methodology

2.1 Enhancement Event Detection

Air pollution concentrations in an urban environment are composed of air originating from different regions affected by different sources. Air from less populated areas upwind of the urban environment, termed the regional background, is transported into the region. As air pollution sources throughout the region are emitted and mix into the atmosphere, an urban background is formed. This urban background is a mixture of pollutants that are either directly emitted or are formed by chemical reactions in the atmosphere. As a result, the complex urban environment has many sources of pollution that vary over different time and length scales. Fundamentally, an enhancement event reflecting a nearby emissions source can be distinguished from the ambient background by differences in spatio-temporal length scales in the sensor time series. Slower variations in the time series (on the order of hours) are attributed to the ambient background while faster variations (on the order of second to minutes) are attributed to local enhancements (Wang et al, 2018, Zimmerman et al., 2019). Within the context of mobile monitoring, the local enhancements may be experienced at even faster timescales as vehicles can travel through emissions plumes within a few seconds. The magnitude of the enhancement will depend on the pollutant, the strength of the source, and the atmospheric conditions, and can range from just above the signal to noise ratio to several orders of magnitude higher than the baseline. This enhancement-based approach has been demonstrated using mobile platforms for a variety of sources including thermogenic and biogenic methane (Weller et al., 2018, Moore et al, 2023).

Figure 1 illustrates some key concepts and definitions used in the description of the data processing methods for detecting enhancement events. While individual implementations of these methods vary for the different data products supporting different emission source types, this section describes the general approach that is consistent across all of the implementations expected for SMMI. The identification of an enhancement event starts with the 1 Hz time series data from individual mobile platforms and generally consists of the following processing steps: 1) deriving the baseline time series from the original time series (Figure 1a), 2) subtracting the baseline time series from the original time series to give a baseline-adjusted time series (Figure 1b), 3) finding enhancement events in the baseline-adjusted time series that are statistically significant, and 4) calculating descriptive statistics about the enhancement event. Figure 1a shows the original time series for a specific pollutant (a black carbon enhancement event is shown, but the example is meant to be illustrative of any pollutant), the derived baseline, and the derived baseline with a constant value added that corresponds to the limit of detection (LOD) of the sensor which we define as three times the standard deviation (σ) of the sensor 1 Hz measurements when measuring a relatively constant concentration (see details in Mobile

Ambient Air Pollution Measurement Quality Assurance System) and denoted as 3σ. A single large enhancement event is highlighted in this example, but a number of smaller enhancement events can also be seen that are slightly higher than the dashed line representing the LOD. Figure 1b shows the baseline-adjusted time series that results from subtracting the baseline from the original time series, with the values representing the enhancement concentration. In this transformed time series, the LOD is represented by a single value, rather than a value that varies in time with the baseline. In situations where it is desired to focus only on the most intense enhancement events, a secondary threshold that is higher than the LOD may be used, as shown in this hypothetical example. The peak enhancement value (illustrated in Figure 1b) is a typical metric used to describe each enhancement event, and is simply defined as the maximum concentration of the baseline adjusted time series during the enhancement event window.



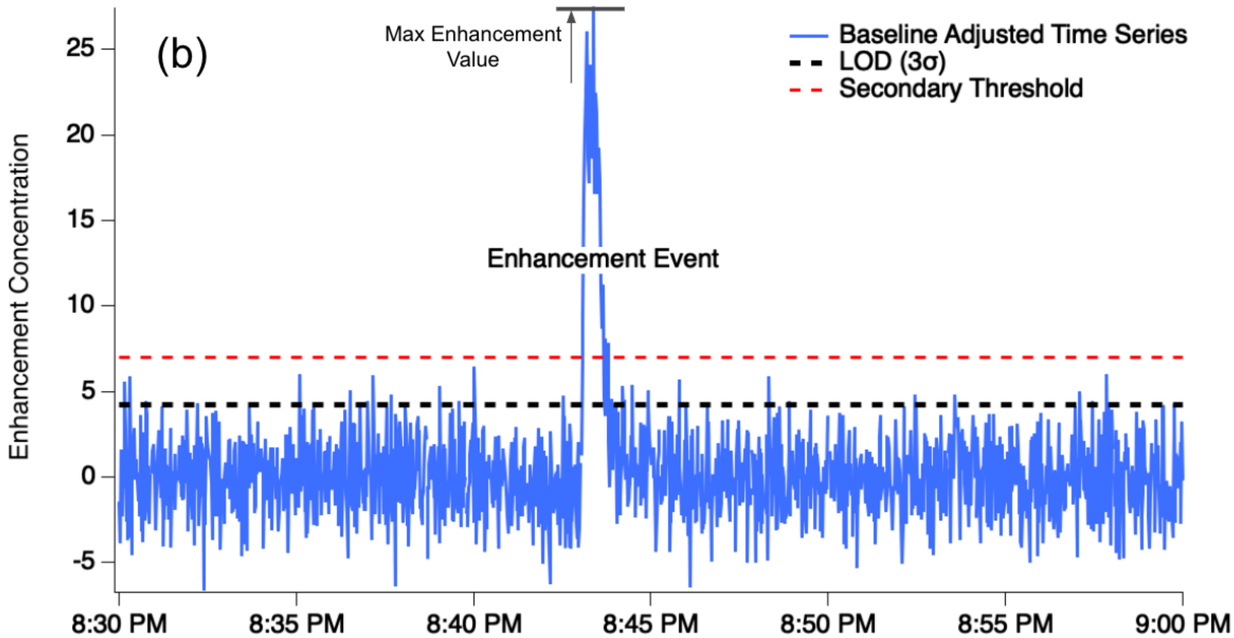


Figure 1: Time series illustrating an example of an enhancement event. Top (a) shows the original time series, the derived baseline calculated using a rolling 30-second median, and the baseline plus the 3**σ** LOD. Bottom (b) shows the baseline adjusted time series calculated by subtracting the baseline from the original time series, with the 3**σ** LOD as well as a hypothetical secondary threshold that could be used in cases where it is desirable to filter out smaller enhancement features that are only slightly above the LOD. Black carbon data is shown, but concepts apply to any pollutant.

The following sections provide additional detail on each of these steps and the different variations used in Aclima's enhancement-based data products.

2.2 Enhancement Event Methodology

2.2.1 Deriving the Baseline

The first step in the data processing pipeline is to derive the baseline time series from the 1 Hz time series. The time series baseline can be derived using any method that isolates the high frequency features from the lower frequency features. In addition to a simple running median approach, some commonly used methods from atmospheric science literature that isolate different frequency components of pollutant time series data include NOAA's CCGCRV model (Thoning et al., 1989) and Kolmogorov–Zurbenko filtering (Wise and Comrie, 2005). Aclima primarily uses running medians for the purpose of deriving the baseline from the original 1 Hz time series.

Regardless of the method used, one of the key input parameters is a temporal length scale that dictates the time/frequency cutoff for inclusion as part of the baseline. The choice of temporal length scale will influence the types of features that are interpreted as enhancements and those that are interpreted as background in a given data product, and the use cases should take this into consideration. Once the baseline is extracted from the 1 Hz time series, the baseline-adjusted time series (shown in Figure 1b) is generated by subtracting the baseline from the original 1 Hz time series. This baseline-adjusted time series is used as the input data for finding enhancement events.

2.3.2 Enhancement Event Identification

We use two general approaches for identifying enhancement events: 1) peak-based events and 2) pass-based events. Each method has advantages and disadvantages, but the primary difference between the two is the spatial scale that is represented most effectively.

Peak-based events are identified using a peak detection algorithm that finds and characterizes peaks in the baseline adjusted time series. The peaks are characterized based on their height (or prominence) and width, their location (latitude and longitude) according to where the maximum of the peak occurs, and the peak start and end times. These characteristics provide some specificity to what a valid enhancement event looks like and can be advantageous for honing in on emissions sources of a specific spatial extent. For example, natural gas leaks from distribution pipelines are typically (but not always) very discrete plumes spatially and it is useful to set minimum and maximum peak widths in order to reduce the influence of other sources and to specifically pinpoint the location where the plume intensity is the highest in order to identify the most likely location for leak repair.

Pass-based events are identified based on an assessment of a relevant statistical property of a drive pass, such as the mean, median, or maximum exceeds some threshold. A drive pass is defined as an aggregation of 1 Hz data collected as a vehicle travels through the defined spatial unit (e.g. road segment, hexbin, etc). In contrast to the peak-based approach, this approach is more flexible and permissive to the characteristics of the enhancement event (i.e. it is not sensitive to the width of the peak), but rather allows for the detection of an event that spans a wider distance than what would be detected using the peak-based approach. A disadvantage is that it will associate a discrete point source event across the entire area covered by the spatial unit; however, it will also more accurately associate the more spatially-distributed enhancement events across the full path that it was detected, which is desirable for capturing the spatial extent from large area source emissions such as certain landfills or farms.

The key factor driving the decision as to what approach is most appropriate for particular data products depends on whether the expected sources are localized discrete sources that are fixed in space vs larger area sources or mobile sources. For localized discrete sources, we use peak-based enhancements and for larger area or mobile sources, we use pass-based enhancements. There is a high degree of interchangeability between the two approaches and while peak-based enhancements might favor more localized events, they will still capture larger area sources, and vice-versa for pass-based enhancements. Understanding the differences between the two approaches is important in interpreting results and understanding limitations.

2.2.3 Enhancement Event Processing

The baseline-adjusted time series is used as the input data set for processing the enhancement events from either of these two methods. Enhancement events are derived from each vehicle separately and subsequently combined into a single dataset that is a compilation of all enhancement events. Each enhancement event carries additional metadata required for tracking and for further data processing downstream including: a unique event id, vehicle id, start and end timestamps, location (latitude, longitude, and where relevant, a spatial unit id), vehicle speed, and vehicle direction. Descriptive metrics about the enhancement events are also calculated and tracked. The key metrics for peak-based enhancements are peak enhancement concentration and peak width (both in units of time and in space). For pass-based enhancements, the key metric is the pass-mean enhancement concentration. Finally, each enhancement event detected for the primary pollutant is associated with information about concurrently detected pollutants in the form of a boolean (true/false) field on whether a particular pollutant enhancement is detected concurrently or not, as well as quantitative metrics describing the magnitude of these concurrent enhancements (if present).

2.2.4 Summary of Critical Input Parameters

The resultant enhancement event data set contains all of the relevant information needed to identify “qualifying” events for being included in the relevant data analysis. (Not all parameters are used for all analyses.) The following is a summary of the relevant input parameters:

- **Sensor limit of detection (LOD)** This is the minimum threshold for enhancement event detection
- **Secondary threshold for enhancement magnitude** In some cases it is desirable to use a secondary threshold for defining qualifying events that may be higher than the sensor LOD
- **Baseline temporal window size or other parameters** At this time, most of Aclima’s data products use a simple running median which requires a single input parameter that defines the window length.
- **Peak-finding input parameters** For peak-based enhancements, a series of input parameters are required defining the search window size, the minimum prominence (difference between the highest and lowest points in the time series), minimum width, and minimum distance between peaks
- **Multi-pollutant criteria** The above parameters also need to be defined for any relevant secondary species that are used in the criteria for identifying source type or categorization. Additionally, the criteria for using enhancements of multiple pollutants to define a specific source type needs to be specified.
- **Vehicle speed filter** Minimum and/or maximum vehicle speed can be defined to ensure that all enhancements are sampled in a relatively consistent manner.
- **Peak width** In some cases the criteria for acceptable peak width range (as both a duration and a distance) is specified in order to ensure that sensor or other artifacts don’t impact the data product.

2.3 Enhancement Clusters

Individual enhancement events provide information about emission sources present in a specific location and at a specific point in time. In order to improve confidence that these enhancement events are tied to a persistent source of emissions in that specific location, the enhancement events detected over time at the same location need to be grouped and characterized as a whole. These groupings of enhancement events will be referred to as enhancement clusters. In order to achieve this clustering in an effective way, some assumptions are required regarding the distance over which two different enhancement events are considered to be from the same source. There are two general approaches: 1) using simple spatial aggregation at a fixed resolution or 2) using a more complex density-based clustering algorithm (i.e. DBSCAN) that can result in enhancement groupings of varying size depending on the spatial extent of enhancement event detections that appear to be associated with the same source.

The advantage of the density-based approach to defining enhancement clusters is that sources which are more broadly spatially distributed may be more accurately characterized by a single cluster of enhancement events, while at the same time maintaining accuracy in the location of clusters with events that are detected close together. This approach works well for data products that represent discrete emissions sources that have limited overlap spatially. The density-based approach does not work well in cases where enhancement events are spatially ubiquitous, such as TVOCs. In these scenarios, entire regions can end up being grouped into a single cluster if the inputs of the enhancement detection and clustering algorithm are not finely tuned to each specific region, which is an approach that does not work well when trying to apply a consistent model across multiple unique collection areas (e.g. urban, rural, suburban etc).

Using a fixed resolution spatial aggregation for defining enhancement clusters is a simple solution for scenarios where the density-based algorithm does not work well. The disadvantage is that assumptions about relevant spatial scales for emission source impact zones are fixed. In some cases multiple distinct hyperlocal sources may get grouped together into a single cluster while larger area sources may get distributed across multiple clusters.
These enhancement clusters are described using a set of descriptive metrics, which vary slightly across the different data products but are generally grouped into two different categories: magnitude metrics and persistence metrics. Magnitude metrics describe the aggregation of enhancement event concentrations. Persistence metrics describe the detection rate of an enhancement event. For magnitude metrics, a commonly used metric is the maximum enhancement value, which is the highest enhancement peak concentration detected within an enhancement cluster. For persistence metrics, some commonly used metrics are peak counts (or number of enhancement events), peak-to-pass ratio (number of peaks detected per pass of a vehicle near the cluster location), and distinct days (number of unique days on which an

enhancement event was detected). For pass-based enhancement events, the same set of persistence metrics apply, but different nomenclature is used: instead of peak counts and peak-to-pass ratio, we use hit counts and hit rate, respectively. Details on specific metrics used are provided within the description of each data product.

The following is a summary of the critical input parameters required for defining the properties of enhancement clusters:

- **Spatial aggregation type and size** Specify either density-based or fixed resolution along with the relevant parameters for the specified type. Density-based aggregation requires multiple parameters, in particular the maximum distance between events. For fixed resolution, the geometry type and size needs to be specified (i.e. hexbin cell level, road segment length, etc.)
- **Minimum number of events or distinct days** In some cases a minimum number of enhancement events needs to be identified in order for a valid cluster to be generated. In addition, the events need to be detected on a minimum number of distinct days. This provides a minimum persistence threshold for the tracking of an event and reduces the number of clusters due to sporadic events or measurement artifacts where it is desirable to have a low false positive rate.

3.0 Impact of Sensor Uncertainty

The Aclima Ambient Air Pollution Measurement Quality Assurance document discusses sensor performance, sources of uncertainty, and QA processes around the mobile platform devices. The enhancement-based data products are not as sensitive to certain sources of uncertainty as the ambient concentration data product. Table 1 summarizes the typical sources of sensor uncertainty and their respective influence on the quality of enhancement-based data products

Table 1. Summary of measurement uncertainties and impacts on enhancement based data products

Uncertainty Source	Enhancement Data Product Sensitivity
Offset Accuracy/Drift	Very Low
Gain Accuracy/Drift	Low
High frequency random noise	High
Anomalous spikes, glitches, fast offset changes, etc	Medium
Interfering Species	Potentially High

Most air quality sensors have different components of uncertainty that vary over different temporal scales. Typically, sensors have high frequency random noise as well as lower frequency variability (often referred to as “drift”). The high frequency noise is often gaussian ("white") noise that is straightforward to characterize and address using standard statistical approaches. For example, any given enhancement event can be quantified by its signal-to-noise value, where the peak enhancement is the signal and the 1σ precision is the noise. If the noise profile is gaussian, then an enhancement with a signal to noise ratio of 3 or greater (i.e. the signal is higher than the 3σ precision above the measurement baseline) has a 99.7% probability of being a significant event and not random measurement error. (For more discussion of the precision of Aclima's sensor suite, see Section 5.5 in the Mobile Ambient Air Pollution Measurement Quality Assurance System.)

Lower frequency sensor errors, or drift, can be harder to predict, vary over different time scales, and vary in magnitude. Drift can be compensated for to some degree as sensors are recalibrated, but in some cases it can become significant in between calibration events. However, if the temporal scale of the ambient background is chosen carefully, this lower frequency error can be lumped in with the baseline, and therefore largely ignored. This is an especially advantageous aspect of these data products within the context of large-scale fleets that are deployed remotely, where it is not practical to perform high-frequency quality control checks of sensor offset and gain to the degree that would typically occur at regulatory monitoring sites. It is important to understand the temporal characteristics of sensor uncertainty and define the time scale for identifying the baseline appropriately or ensure that the lower limit of the enhancement magnitude is set high enough to rule out drift as the source of any enhancement. Aclima currently uses very short (minutes or less) length scales for defining baselines in all enhancement-based data products, which effectively results in drift being a negligible source of uncertainty in most cases. Sensor uncertainty at faster time scales is assumed to be purely gaussian and is characterized by the sensor LOD with a specified confidence level (i.e. 2σ, 3σ, etc) and used as the minimum threshold for identifying enhancement events.

The above description holds if the sensor drift is primarily in the offset of the sensor, i.e. it is a fixed value regardless of the measured concentration. In this scenario, the enhancement magnitude will not be impacted because the drift is equivalent for the baseline and the maximum value of the enhancement event; thus, the difference remains unaffected. However, if the drift is in the sensor gain (i.e. the error scales with concentration) the accuracy of the magnitude of detected enhancement events will be impacted equivalently. Both gain and offset uncertainty (and drift over time) are characterized in Aclima's post-deployment calibration process, as described in the Mobile Ambient Air Pollution Measurement Platform Quality Assurance System document. Table 2 shows typical gain uncertainty for the relevant measurements used in Aclima's enhancement-based data products. These uncertainties in gain are derived from calibration checks relative to their prior calibrations, representing drifts over 3-12 months. These uncertainties propagate directly to the measurement uncertainty for individual enhancement events, but in general these uncertainties are very small compared to variability in the behavior of an emissions plume. For example, Von Fischer et al. (2017) and Weller et al. (2019) show significant variability in observed concentrations and estimated leak rates from repeated sampling of the same leak . Variability and uncertainty for real world emissions sources in complex urban areas is expected to be even higher. Sensor gain uncertainty, therefore, is expected to be negligible within this context.

Table 2. Typical Sensor Gain Uncertainties for Aclima Mobile Platform, derived from calibration checks

Pollutant	**Typical Gain Uncertainty**
Methane	1%
Ethane	1%
TVOC	32%[1]
Blackcarbon	3%
NO	22%
CO	6%
PM2.5	10%

[1] Sensitivity of TVOC sensor is highly variable for different classes of organic compounds.

An additional source of uncertainty has to do with randomly occurring anomalous sensor artifacts such as noise spikes, fast offset changes, glitches, etc. These are not common but can occasionally occur, especially when a sensor malfunctions. As documented in the Mobile Ambient Air Pollution Measurement Quality Assurance System document, many of these types of events are detected and flagged automatically as part of the data processing pipeline. However, we do also discover some manually in our data review process, which are flagged for omission after the fact, prior to final data verification. In certain cases, depending on the nature of the artifact, these events can register as individual enhancement events while mapping is ongoing. Eventually, these events will get flagged and removed by the data review team. However, since these data products may trigger intervention before data is fully verified, it's important to consider their impact. As discussed in the introduction, there are two features of these data products that can minimize the impact of these artifacts: 1) multipass sampling and 2) multiple pollutant inputs.

Leveraging multiple passes in these data products builds confidence in a cluster of enhancement events, under the assumption that it would be highly unlikely for one of these artifacts to occur in the same exact location multiple times. In this sense, the persistence metrics can be a strong indication of confidence in a particular source type in a particular location. Multiple pollutant inputs as criteria for identifying qualifying enhancement events also reduces the probability of detecting false positives due to sensor artifacts (spikes, glitches, etc) because it is unlikely that two sensors operating on two different detection principles will exhibit random artifacts at the same time. The more specific the criteria are for identifying an enhancement, the less likely it is for an artifact to result in a false positive. For example, including an acceptable range of ratios between two pollutants will be more selective than the criteria that two pollutants must undergo enhancements simultaneously.

4.0 General Use Cases and Limitations

4.1 Use Cases

In general, these data products identify locations where specified sources of emissions are occurring and offer insights into the most effective ways to intervene or mitigate these sources. There are two general data product use cases:

1) Identify the locations of specific emissions sources where a direct intervention can lead to a reduction or an end to the individual emissions source (e.g. natural gas leak detection to flag for repair); the data products can additionally provide a means to rank the severity or consistency of the emissions source in order to prioritize the interventions.
2) Analyze the spatial distribution of different source types indicated by the data products to better understand which general activities might correlate with a higher density of enhancement clusters (e.g. landfills and biogenic methane; TVOCs and gas stations, etc.).

As we have described, enhancement-based data products are only minimally sensitive to sensor drift and other anomalous sensor artifacts. As such, enhancement-based data products can provide value for users in the early stages of data collection in a given region, and actionable insights can be gleaned from the data without waiting for data to enter the Verified data stage (see Mobile Aclima Ambient Air Pollution Measurement Quality Assurance System document for more details about data stages). This is particularly true for use case 1 outlined above — identifying individual emission sources that may emerge as data collection is on-going. For use case 2, additional quality assurance must be worked into the analysis by the end user in order to ensure relatively equal and complete sampling for comparing areas of high and low enhancement cluster densities. As a result, use case 2 may not be as effective until all collection is complete. However, a major advantage of Aclima's enhancement-based data products is that there is a lower likelihood of detecting false positives and this is true even while data are in the Preliminary data stage.

4.2 Limitations

Specific implementations of enhancement-based data products each have their own limitations that should be considered when interpreting results:

- **Not ideal for catching sporadic emissions sources.** Because of the nature of mobile data collection, a highly sporadic emission source may not be detected; more persistent or continuous sources of emissions have a much higher probability of being detected.
- **Variations in wind direction make precise source location identification challenging.** While wind data can be used to improve location identification, there may still be challenges to precisely identifying source locations and the location the enhancements are detected may not reflect the location of the responsible source. Therefore, caution and careful consideration should be taken before assignment of observed enhancement features to specific sources, especially for sources not located immediately on-road or roadside.
- **Emission rates may not scale directly with the enhancement concentration**. Quantitative properties of the peaks and clusters can be used to prioritize certain emissions sources for mitigation or further investigation, but these metrics should be interpreted as only qualitative or, at best, semi-quantitative and cannot be used to directly infer emission rates. In some cases, the sensor may also be a large source of uncertainty in the calculation of the concentrations present during an enhancement event (e.g. TVOCs). However, even in cases where the true concentrations are accurately measured, plume dispersion dynamics are a large source of uncertainty in estimating emission rates from concentrations.
- **Sensitivity to time window for baseline definition.** The spatial extent of individual emissions sources may not be accurately captured in cases where the time it takes to transect the plume is on the same order or longer than the temporal length scale of the baseline derivation. We expect different sources of interest to have different spatio-temporal length scales. For example, a discrete point source located at street level such as an idling bus may have an associated emissions plume that is a few meters wide and can be driven through in a few seconds, whereas a large area source such as a landfill or a farm may have distributed emissions plume over hundreds or even thousands of meters, which would take longer to pass through in a mobile platform. This translates to the need to properly select the time scale used to isolate the higher frequency components of the time series.

5.0 Citations

Moore, D.P., Li, N.P., Wendt, L.P., Castañeda, S.R., Falinski, M.M., Zhu, J., Song, C., Ren, Z.J., and Zondlo, M.A. (2023) Underestimation of sector-wide methane emissions from United States wastewater treatment, *Environ. Sci. Technol. 57* (10), 4082-4090, doi: 10.1021/acs.est.2c05373

Thoning, K. W., Tans, P. P., and Komhyr, W. D. (1989) Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985, *J. Geophys. Res.*, 94(D6), 8549–8565, doi:10.1029/JD094iD06p08549.

von Fischer, J.C., Cooley, D., Chamberlain, S., Gaylord, A., Griebenow, C.J., Hamburg, S.P., Salo, J., Schumacher, R., Theobald, D., and Ham, J. (2017) Rapid, vehicle-based identification of location and magnitude of urban natural gas pipeline leaks, *Environ.l Sci. Technol.* 51 (7), 4091-4099, doi: 10.1021/acs.est.6b06095

Wang, J.M., Jeong, C-H., Hilker, N., Shairsingh, K.K., Healy, R.M., Sofowote, U., Debosz, J., Su, Y., McGaughey, M., Doerksen, G., Munoz, T., White, L., Herod, D., and Evans, G.J. (2018) Near-road air pollutant measurements: Accounting for inter-site variability using emission factors *Environ. Sci. Technol.* 52, 16, 9495–9504, doi:10.1021/acs.est.8b01914

Weller, Z.D., Roscioli, J.R., Daube, W.C., Lamb, B.K., Ferrara, T.W., Brewer, P.E., and von Fischer, J.C. (2018) Vehicle-based methane surveys for finding natural gas leaks and estimating their size: validation and uncertainty, *Environ. Sci. Technol. 52* (20), 11922-11930, doi: 10.1021/acs.est.8b03135

Wise, E.K., and Comrie, A.C. (2005) Extending the Kolmogorov–Zurbenko Filter: Application to Ozone, Particulate Matter, and Meteorological Trends, Journal of the Air & Waste Management Association, 55:8, 1208-1216, doi 10.1080/10473289.2005.10464718

Zimmerman, N., Li, H.Z., Ellis, A., Hauryliuk, A., Robinson, E.S., Gu, P., Shah, R.U., Ye, Q., Snell, L., Subramanian, R., Robinson, A.L., Apte, J.S. and Presto, A.A. (2020). Improving Correlations between Land Use and Air Pollutant Concentrations Using Wavelet Analysis: Insights from a Low-cost Sensor Network. *Aerosol Air Qual. Res.* 20: 314-328. doi:10.4209/aaqr.2019.03.0124

7.0 Appendix

7.1 Definitions

Distinct Days Count: The number of unique days on which an enhancement peak or event was detected in a specified location.

Drive pass: defined as an aggregation of 1 Hz data collected as a vehicle travels through a defined spatial unit (road segment, hexbin, etc.)

Enhancement clusters: Spatial groupings of individual enhancement events detected at different times, but within the same location.

Enhancement events: Localized elevation in concentration of a given pollutant within an emission plume that is measurably distinguishable from the ambient background.

Limit of detection (LOD): Defined as three times the standard deviation (σ) of the sensor 1 Hz measurements when measuring a relatively constant concentration (see details in Mobile Ambient Air Pollution Measurement Quality Assurance System) and denoted as 3σ.

Magnitude metrics: A class of metrics that describes the magnitude of a specific enhancement event (as a concentration) or an aggregation of the magnitude of multiple peaks detected within the same enhancement cluster. Peak enhancement is an example of an enhancement event magnitude metric. Max enhancement is an example of an aggregated enhancement cluster magnitude metric (i.e. the maximum peak enhancement across all peaks associated with a given cluster)

Pass-based enhancement event: Enhancement events that are identified based on an assessment of a relevant statistical property of a drive pass, such as the mean, median, or maximum exceeds some threshold.

Peak-based enhancement event: Enhancement events that are identified using a peak detection algorithm that finds and characterizes peaks in the baseline adjusted time series.

Peak-to-pass ratio: Describes the percentage of times where an enhancement event of a particular type is detected for all passes of the spatial unit.

Persistence Metrics: A class of metrics that describes the consistency over time with which a specific location is impacted by a particular emissions source. Peak-to-pass ratio is an example of persistence metrics.

Regional background: Air from less populated areas upwind of the urban environment.

Urban background: As air pollution sources throughout the region are emitted and mix into the atmosphere, an urban background is formed, which is a slowly-varying mixture of pollutants that are either directly emitted or are formed by chemical reactions in the atmosphere.

