HHS Public Access

Author manuscript

Atmos Environ (1994). Author manuscript; available in PMC 2021 July 15.

Published in final edited form as:

Atmos Environ (1994). 2020 July 15; 233: . doi:10.1016/j.atmosenv.2020.117519.

Using Gas-Phase Air Quality Sensors to Disentangle Potential Sources in a Los Angeles Neighborhood

Ashley Collier-Oxandale[1], **Nicole Wong**[2], **Sandy Navarro**[3], **Jill Johnston**[4], **Michael Hannigan**[5]

[1]Environmental Engineering, University of Colorado Boulder

[2]Redeemer Community Partnership

[3]Esperanza Community Housing

[4]Keck School of Medicine, University of Southern California

[5]Mechanical Engineering University of Colorado Boulder

Abstract

In the late summer of 2016, our team deployed a network of low-cost air quality sensing systems in partnership with community-based organizations in a neighborhood in South Los Angeles, California. Residents of this community were concerned about possible emissions from local oil and gas activity, however in addition to these potential emissions, the neighborhood is also subject to a complex mixture of pollutants from other nearby sources including major highways. For this deployment, metal-oxide VOC sensors were quantified to provide methane (CH_4) and total non-methane hydrocarbon (TNMHCs) concentration estimates. This data along with other sensor signals, meteorological data, and community member observations was used to examine the composition and possible origins of observed emissions. The sensor network displayed expected environmental trends and highlighted short-term elevations in CH_4 and/or TNMHCs, which we were then able to investigate more closely. The results indicated that sources of both combusted and volatilized hydrocarbons were likely affecting air quality throughout the community, including near the site of the local oil and gas activity. This deployment may serve as a model for how multi-sensor systems deployed in networks can be leveraged to better understand sources in complex areas, potentially supporting future community-based air quality research.

Keywords

low-cost sensors; gas-phase sensors; air quality; multi-sensor; VOC sensors

Note 1: (regarding reference air quality data) this data has not passed through the normal review process and is therefore not QA/QC'd and is unofficial data. Note 2: (regarding the first author's current position with the South Coast Air Quality Management District): this is a research article and does not constitute an endorsement or recommendation of a particular low-cost sensor or low-cost sensor platform.

1. INTRODUCTION

The implementation of the Clean Air Act has led to immense improvements in air quality, particularly at the regional scale. However, the benefits of air quality improvements have not been realized equally across communities and more attention could be focused on variability in air pollution at the local level (Miranda et al., 2011). Monitoring near roadways (~2 to 400 m) consistently observes steep pollutant gradients near high-traffic roads (Brugge et al., 2007) and residents living in close proximity may face elevated health risks. Concentrations of benzene and summed volatile organic compounds (VOCs) have been found to be twice as high outside of homes in areas with heavy traffic as compared to those with low traffic (Fischer et al., 2000). Sources other than roadways can drive this small-scale variability as well. In addition to traffic, residential ambient concentrations of VOCs have been measured at levels 1.5–4 times higher within 50 meters of an industrial source (e.g., petroleum hydrocarbons or BTEX compounds near gas stations or perchloroethylene near dry cleaners) (Kwon et al., 2006). Significantly higher concentrations of BTEX compounds and styrene have been measured on the fenceline of a refinery as compared to sites approximately a mile away (Mukerjee et al., 2016). Given the potential health risks posed by some of these compounds emitted by large and small-scale industries (Loh et. al., 2007), a more comprehensive understanding of the local air quality and the variability of pollutants within communities could inform actions to improve public health. More detail at the local level might be particularly valuable in high-density locations with many potential sources.

1.1 Motivation for Community-driven Air Monitoring in South Los Angeles

While Los Angeles, California (CA) has long been the subject of research and policies to address poor air quality resulting from numerous point and mobile emission sources, continued growth, topography, and meteorology (Kunzli et al., 2003), neighborhood level air quality measurements have been less well characterized. One potential source of air pollutants of concern to communities is urban oil and gas extraction. Los Angeles County is home to more than 5000 active and idle oil wells, with 850 of these located within the City of Los Angeles (Sadd & Shamasunder, 2015). Rapid development over the years has led to a “conflict in land usage” (Chilingar et al., 2005), which has resulted in the active wells and production sites in close proximity to high density residential areas and public services such as schools and hospitals (Shamasunder et al., 2018). Given the potential for the release of air toxics from these facilities, such as BTEX (benzene, toluene, ethylbenzene, and xylene) compounds (Adgate et al., 2014; Helmig et al., 2014; Moore et al., 2014), increased near-source monitoring could help to characterize the impact on local air quality. Motivated by a concern for their health, residents of these communities often seek ways to collect more information or data to better understand their exposure and potentially facilitate action to reduce exposure (Brown, 1992). The increasing availability of more accessible environmental monitoring technologies, coupled with greater engagement, could support local community-based air quality research projects.

1.2 Background on Low-Cost Air Quality Sensors

One of these increasingly accessible technologies is low-cost air quality sensors, which have the potential to provide preliminary data to inform more targeted studies, to supplement

existing monitoring networks, and to aid in the quicker detection of pollution hotspots (Snyder et al., 2013). The cost of these systems typically ranges from approximately $500 - $5000 each making the deployment of networks of sensor systems more feasible. Furthermore, because of the lower costs and the relatively simple deployment and operation procedures, this technology is well-suited to support community-based investigations (Shamasunder et al., 2018). However, an ongoing challenge associated with the use of low-cost sensors is ensuring high quality data. These sensors exhibit cross-sensitivities to environmental parameters and confounding pollutants (Lewis et al., 2016). Significant research, both in the lab and the field, has been undertaken to better understand and mitigate these cross-sensitivities as well as to develop different approaches to calibration methods and models (De Vito et al., 2008; Eugster & Kling, 2012; Mead et al., 2013; Piedrahita et al., 2014; Leidinger et al., 2014; Masson et al., 2015a & 2015b; Sadighi et al., 2018; Hagan et al., 2018; Kizel et al., 2018; Collier-Oxandale et al., 2018a; Collier-Oxandale et al., 2019). Beyond sensor quantification, researchers have explored the use of sensors in various applications such as studying the spatial variability of pollutants (Bart et al., 2014; Moltchanov et al., 2014; Sadighi et al., 2018; Cheadle et al., 2017), examining indoor air quality (Casey et al., 2018), to better understand specific point sources (Thoma et al., 2016; Yuval et al., 2019), and to support personal exposure monitoring (Jerrett et al., 2017, English et al., 2016).

Using low-cost VOC sensors can be particularly challenging given their typical lack of selectivity and susceptibility to interferents (Spinelle et al., 2017). However, there are studies illustrating the use of VOC sensors to detect methane (CH_4) at ambient levels as well as studies where unique techniques (e.g., multi-sensor arrays or temperature-controlled operation) have been used to identify target VOCS in the presence of confounders (Eugster & Kling, 2012; Leidinger et al., 2014). Studies have also explored the use of the sensors in complex ambient environments (Collier-Oxandale et al., 2018a & 2019), and how machine learning may be used with data from sensor arrays to identify source types (Thorson et al., 2019). As quantification approaches for VOC sensors continue to improve and best practices are established it is possible that these sensors could play a valuable role, alongside other types of low-cost air quality sensors, in addressing environmental inequities by helping to identify communities or neighborhoods overburdened by air pollution.

1.3 This Study

In this paper, we apply previously demonstrated approaches to estimate methane (CH_4) and total non-methane hydrocarbon (TNMHC) concentrations using measurements from a low-cost sensor network (Collier-Oxandale et al., 2018a & 2019). These estimates are then utilized along with supplementary data to study air quality trends and identify potential emission events at the local level in a South Los Angeles neighborhood. This project was conducted in a participatory manner and in partnership with two local, community-based organizations, Redeemer Community Partnership and Esperanza Community Housing. The deployment spanned primarily two neighborhoods in South Los Angeles, West Adams and North University Park. In both communities, problems such as poverty and housing insecurity contribute to community vulnerabilities. West Adams is made up of 87% residents of color, including 58% Latino and 20% African American. Furthermore 68% of residents

live at or below 200% the poverty line (Shamasunder et al., 2018). University Park is predominantly Latino at 76% and here 72% of residents are living at or below 200% the poverty line (Shamasunder et al., 2018). Prior studies have demonstrated the reality of environmental injustices resulting in higher levels of exposure to air pollutants in minority and socioeconomically disadvantaged communities (Souza et al., 2009; Marshall, 2008; Clark et al., 2017). In some cases, odors or the visibility of industrial activities (e.g. acidization or flaring) may also contribute to increased stress or adverse effects on residents' quality of life (Beloff et al., 2000). Driven by community concerns, the two partner organizations brought awareness to potential adverse impacts of neighborhood oil drilling on residentials health and environmental quality (Sadd & Shamasunder, 2015). Informed by community concerns, we piloted an approach using low-cost sensors in a complex urban environment with the goal of learning more about local sources. Furthermore, the use of a multi-sensor platform along with the various analysis techniques shared in this paper may serve as a model for future community-based air quality research projects.

2. METHODS

2.1 Deployment Overview

Fifteen low-cost sensing systems were deployed for a period of 8 weeks, from late summer through early fall 2016, in a community in South Los Angeles. Figure 1 depicts the sampling sites in relation to sources of interest; the sites are distributed on either side of and at varying distances away from the major highways indicated by black lines and an active multi-well oil extraction site indicated by a red star. The community identified the extraction site as being of interest, as a result much of the analysis examines spatial and temporal trends with respect to this location. The sampling sites were based on community interest, recommendations by researchers, and feasibility of access. Fourteen of the sites were located within a roughly 5×5 km area. Sites C1 and C2 were selected as they allowed for continuous co-location with reference instruments providing additional validation data. The neighborhood in which the sensor systems were deployed is primarily high density residential with some commercial and industrial land use. While only the major highways are indicated, some sites were located on high-traffic roads where substantial vehicle emissions can occur as well. While there are other active and inactive oil extraction and processing sites in the area, there were no active drill sites immediately near any sampling site other than the one already noted. Other potential local sources of VOCs would include, most prominently, gas stations, autobody shops, and dry cleaners. While the oil extraction facility is the closest potential source to sites E1 and E2, there is also a dry cleaner to the south of site E1 and an auto repair shop to the southwest of E1. There are several gas stations to the north, near site N5, and to the east, near site R4 and close to the highway. . Several other potential point sources, such as dry cleaners and auto body shops, are located throughout the community and on all sides of the two surrounding highways.

Placement of the systems at each site was primarily based on safety, convenience, and resident preference, although factors such as elevation above the ground and optimal access to air flow representative of surrounding conditions were also considered. As observed in a previous study, intra-site variability can occur and may be driven by sources in close

proximity to the site that the sensors may or may not be well-positioned to detect (Collier-Oxandale et al., 2018b). For this reason, we ensured that the two Y-Pods adjacent to the site of interest had a “line-of-sight” to the potential source.

As previously described, the project utilized community-based participatory research methods. Together we worked with local partners to plan the project, choose the sampling sites, and conduct the sampling with assistance from local residents and businesses who hosted the monitors. Partners also assisted with finding and hiring a local field technician to check the sensor systems periodically and collect the data. As part of the project a Memorandum of Understanding (MOU) was developed and signed by all partners to ensure a mutual understanding of the limitations of sensor technology and study objectives as well as to ensure ongoing communication occurred, particularly around the dissemination of the results of the project. Further, we integrated local knowledge and observations of community members. Following the deployment, we were provided with a log of observed activity at the drill site and records of noise and odor complaints.

2.2 Low-Cost Sensing Systems

This study used an open-source, low-cost sensor system. These systems, termed U-Pods or Y-Pods, include various commercially available sensors, including metal oxide semi-conductor (MOx), electrochemical (EC), and non-dispersive infrared (NDIR) sensors for measuring gas-phase pollutants and other environmental parameters (see Table S1 and Figure S1). U-Pods and Y-Pods are iterations of the same design with some minor differences in the hardware and software intended to increase the reliability of the platform in the updated version; more details are available in previous publications (Masson et al., 2015b; Sadighi et al., 2018; Collier-Oxandale et al., 2018a & 2018b; Casey et al., 2018; Cheadle et al., 2017; Casey & Hannigan., 2018, Collier-Oxandale et al., 2019).

The sensors that this analysis relies on primarily are the Figaro TGS 2600 and the Figaro TGS 2602, both of which are MOx sensors. These sensors are described by the manufacturer as intended for the detection of “air contaminants”, typically used in industrial applications. The datasheets list different compounds each sensor is responsive to, which along with previous research, suggests that the two models exhibit unique selectivities (Figaro, 2015a & 2015b, Collier-Oxandale et al., 2019). The stated detection ranges for both is 1 – 30 ppm (Figaro, 2015a & 2015b), however, this does not take into account the approach of utilizing more advanced signal processing techniques. For example, researchers have explored the potential for the Figaro 2600 to detect CH_4 and the potential for the Figaro 2602 to detect individual, groups of, and total VOCs (Becher et al., 2010; De Vito et al., 2011; Eugster & Kling 2014; Collier-Oxandale et al., 2018a, 2018b, & 2019). An additional study focusing on the variability among multiple Figaro 2602 sensor responses observed relatively low intra-sensor variability over a period of a few hours when sensors were exposed to zero air (< 10 pbb VOCs) (Smith et al., 2017). During a longer period of 20 days high intra-sensor correlation was continually observed, however drift was also observed resulting in an increasing spread among the sensor signals over time (Smith et al., 2017). The authors suggested that deploying clusters of these sensors may mitigate this issue (Smith et al., 2017).

Another important issue is the susceptibility of low-cost air quality sensors to interferents. The performance of MOx sensors may be impacted by environmental conditions in a variety of different ways. For example, temperature can directly impact a sensor's hardware affecting its ability to function, temperature can also govern reaction rates between the metal oxide surface and the target gas in the atmosphere (Wang et al., 2010; Sun et al., 2012). MOx sensors' cross-sensitivity to other gases is especially problematic for low-cost VOC sensors, which tend to be sensitive to both a wide array of VOCs and confounding compounds (Spinelle et al., 2017). As an example, the Figaro TGS 2600 has been observed to be cross-sensitive to carbon monoxide (Collier-Oxandale et al., 2018a); while the Figaro 2602 is sensitive to compounds such as ammonia and hydrogen sulfide, in addition to an array of VOCS (Figaro, 2015b). It is possible that unknown interferents may be impacting the sensors as well.

2.3 Sensor Signal Processing & Quantification

2.3.1 Field Calibration—Given the susceptibility of low-cost sensors to temperature and humidity as well as confounding gases, calibration is a necessary step. While laboratory tests are useful for determining the capabilities of sensors, researchers have continually found that field co-locations allow for the generation of more accurate calibration models (Piedrahita et al., 2014; Castell et al., 2017). The reason for this is likely the large and dynamic range of environmental variables and presence of background pollutants, which cannot be adequately replicated in the lab. This technique involves co-locating sensors with reliable reference instrumentation before and after a field deployment and then using this data along with a technique such as multiple linear regression or machine learning to generate and evaluate a calibration model (Casey at al., 2019; Cross et al., 2017; Zimmerman et al., 2018). Using this method, the field co-location provides data to train the calibration model to detect the target pollutant amid environmental and background conditions that are ideally similar to those experienced during a field deployment.

While field co-locations tend to produce more robust calibration models, there are limitations to this approach. As the calibration model is based on the conditions experienced during the co-location period, if temperature or humidity exceeds the range experienced during co-location, then the model will be extrapolating and may produce less reliable pollutant estimates (Eapoaito et al., 2017). In one study exploring the transferability of calibrations between different locations, researchers observed that calibration models maintain higher performance when they are transferred between areas impacted by similar sources as compared to sites closer together geographically, but impacted by different pollutant sources (Casey & Hannigan, 2018). It has also been observed that high performance models utilizing techniques such as machine learning tend to over-fit to the conditions of the co-location site, which results in diminished performance at new locations (Vikram et al., 2019).

In this study, the selection of co-location sites was based on availability and the support of local regulatory agencies, which allowed for both pre-deployment and post-deployment co-locations with high quality reference instruments (Table 1). The pre-deployment co-location, for CH_4 and TNMHCs occurred at a site in Los Angeles, this site was in a suburban area

that, in general, tended to experience lower levels of the target pollutants. The post-deployment co-location for CH_4 occurred at a mixed-use site in Los Angeles similar to the field deployment sites in terms of land-use. The post-deployment co-location for TNMHCs occurred at a rural site north of Los Angeles in Shafter, CA. This site had some similarities to the field deployment sites in terms of potential nearby sources. While there are important differences to note between the co-location and field deployment sites, the calibration models utilize relatively simple multiple linear regression techniques as opposed to more complex machine learning techniques, which may support enhanced transferability of the models (Vikram et al., 2019).

2.3.2 Signal Processing and Calibration Models—Initial sensor signal processing involves converting the data to a normalized resistance, removing data from the warm-up period, and calculating minute-medians from the raw data. This process is described in greater detail in previous publications (Collier-Oxandale et al., 2018a & 2019). This processed sensor data is then used along with the reference data to generate the calibration models (Table 2). The calibration models rely on data from the Figaro TGS 2600 and Figaro TGS 2602, made by Figaro Inc., along with environmental data (i.e., temperature and humidity) to convert sensor signals to concentrations. In total three models are used, two to estimate CH_4 levels and one to estimate TNMCH levels. The difference between the two CH_4 models is that the first model relies on data from a single VOC sensor while the second model incorporates data from both MOx VOC sensors. The first model was determined to be the best-fitting model in a previous study as it seemed to correct for complex temperature and humidity effects (Collier-Oxandale et al., 2018a), however subsequent work suggests that multi-sensor models may help address cross-sensitivities to some confounding pollutants and provided better overall estimates of target species (Collier-Oxandale et al., 2019). Therefore, both models are included in this analysis. For this study, we used the same model for both the multi-sensor CH_4 and TNMHC models. As demonstrated during previous work, multi-sensor calibration models with the same structure and based on the same sensor data, but different reference data can be specialized to target specific pollutants or groups of pollutants (Collier-Oxandale et al., 2019).

2.3.3 Supporting Data & Additional Processing—Signals from EC carbon monoxide (CO) sensors and NDIR carbon dioxide (CO_2) sensors were also utilized in this analysis to better understand the composition of potential emissions events. In several studies, EC sensors from the same manufacturer have been used to examine spatiotemporal trends and to compare different approaches to sensor calibration (Mead et al., 2013; Cross et al., 2017; Zimmerman et al., 2018; Jerrett et al., 2017). Additionally, studies utilizing the EC CO sensor and the NDIR CO_2 listed in Table S1, have observed uncertainties of 0.1 – 0.44 ppm for CO and ~10 ppm for CO_2 (Piedrahita et al., 2014; Casey et al., 2018; Collier-Oxandale et al., 2018b). For this study, these signals were also quantified using field calibration. While it is important to continue to test the capability of these sensors, the work presented here will focus on the VOC sensor quantification and analysis. Other supplemental data used includes local wind speed and direction data collected using a low-cost anemometer and wind vane located at one of the deployment sites (E2).

In addition to the inclusion of this supplemental data, a signal processing technique developed for atmospheric data was used to identify and remove the baseline from the pollutant data in order to highlight the short-term increases (Ruckstuhl et al., 2012). Heimann and colleagues (2015), demonstrated the application of this technique to low-cost sensor data to separate regional trends (represented by the baseline) from local events (apparent in the baseline-removed data). Figure 2 illustrates the application of this technique to a roughly two weeks of CH_4 data from a single site. In this figure, the diurnal trends are no longer present in the baseline-removed data and the increases that do occur are shorter in duration and do not seem to occur at regular intervals. These increases may represent "events" potentially driven by local air quality conditions or by another phenomena eliciting a response from the sensors. Wavelet analysis is another approach to baseline removal that has been used to study the variability of pollutants on different time scales (Etzion & Broday, 2018).

3. RESULTS & DISCUSSION

3.1 Sensor Performance Quantification

3.1.1 Targeting Methane and Non-methane Hydrocarbons—Table 3 displays the average of the statistics that result from quantifying the 15 sensor systems to estimate CH_4 and TNMHCs. The complete statistics for each individual Y-Pod as well as timeseries of the reference data and fitted data are available in the Supplemental Materials (Tables S2 & S3, Figures S2 – S4). When compared to previous studies using the same sensors in the same platform, the results are similar suggesting there may be some consistency in the behavior of these MOx sensors. For instance, the average R^2 and RMSE for the testing data for a subset of three Y-Pods quantified in a previous study, using the single sensor calibration model for CH_4, (Collier-Oxandale et al., 2018a) were 0.74 and 0.18 ppm respectively. Comparing the performance of the two multi-sensor models in Table 3 to previous work reveals similarities as well. For CH_4, the multi-sensor model consistently exhibits improved performance over the single sensor model (Collier-Oxandale et al., 2018a & 2019). For TNMHCs, the R^2 resulting from using a multi sensor model to estimate a summed VOC signal was 0.59 (Collier-Oxandale et al., 2019).

In this analysis, the estimates resulting from both the single and multi-sensor CH_4 models are considered. While the multi-sensor model appears to improve estimates of CH_4 levels, as indicated by the performance statistics, it is possible that the inclusion of the second VOC sensor in the calibration model may increase the potential for this calibration model to respond to other NMHCs as well as CH_4. As has been previously illustrated, these two VOC sensors exhibit different selectivities, thus including both may widen the range of VOCs that may elicit a response (Collier-Oxandale et al., 2019). Figure S5, available in the Supplemental Materials, highlights the differences between the single and multi-sensor models through their application to additional validation data collected during the field deployment period. Given that the amount of data available to directly compare the single and multi-sensor CH_4 calibration models in relation to both CH_4 and TNMHC reference data were limited (as a result of the available reference instrumentals at the different co-location sites), we were not able to conduct a thorough comparison of the two approaches

for this dataset. However, a better understanding of this issue would be valuable if these sensors continue to be used in the field.

As this study is exploring the potential for these sensors to indicate significant differences over small spatial and temporal scales, differences that may indicate the presence of local sources or phenomena worthy of further investigation, it's important to also consider similarities in the sensor signals. Table 4 displays the average intra-sensor coefficient of determination (R^2) and the average intra-sensor accuracy in the form of the RMSE. This data reveals high agreement amongst the sensors, which has been observed previously; for example, studies examining co-located MOx sensors have observed high correlation (Collier-Oxandale et al., 2018b; Cheadle et al., 2017; Sadighi et al., 2018). Furthermore, researchers studying the performance of eight Figaro TGS 2602 sensors found high intra-sensor correlation and relatively low variability amongst sensors when exposed to zero air, on the order of approximately +/− 5 ppb VOCs (Smith et al., 2017), which is similar to the average accuracy observed here for TNMHCs. Though, it is important to note that this test was performed over several hours and the results may vary over longer time scales of days or weeks.

Table 4 also provides the signal to noise ratio (S/N) for the co-location period calculated as the median over the standard deviation (Piedrahita et al., 2014). Comparing the S/N values for the CH_4 estimates from low-cost sensor systems to the S/N values for the CH_4 reference data further emphasizes the agreement between the reference data and the fitted sensor data. It is more difficult to draw conclusions regarding the TNMHC S/N values, as the trends appear to be inherently noisier given that the S/N value for the reference data that is less than 1.0. Though when examining the timeseries data, excerpts from two separate days are shown in Figure 3, there are instances where increases in the fitted sensor signal coincides with increases in the reference data. There are also instances where an increase occurs in either the reference data or fitted sensor data that is not reflected by the other, however, it's worth noting that when increases occur in the fitted sensor data they are consistently reflected across all of the Y-Pods. The agreement in these trends suggests that what is driving the increases must be environmental as opposed to random noise, such as exposure to a target pollutant, a confounding compound, and/or a change in environmental conditions.

In terms of data completeness, there were periods of continuous data loss for two Y-Pods that do not appear to have impacted data quality, thus the available data from these Y-Pods has been included in the analysis. However, another a third Y-Pod experienced repeated, sporadic power loss resulting in unreliable data, and this Y-Pod has been excluded from the analysis. A fourth Y-Pod excluded from the analysis appears to have had a problem with a VOC sensor. The calibration model results utilizing data from this sensor were significantly different from the other results and this sensor exhibited poor correlation with other sensors of the same model. Thus, data from the remaining 14 Y-Pods were utilized in the analysis. Supplementary data from U-Pods co-located at certain field sites provided additional streams of data and that has also been included.

3.1.2 Additional Gas-Phase Sensor Quantification Results—In addition to CH_4 and TNMHCs, the CO_2 and CO sensor signals were also quantified. Table 5 shows the

summary statistics with the number of Y-Pods included. There are fewer CO sensors quantified as these sensors were not integrated into or co-located with all deployed Y-Pods. These uncertainties are similar to the results of other studies utilizing the same sensors in the same or similar platforms. For example, RMSEs of 10.1 ppm (Collier-Oxandale et al., 2018b) and standard errors of 9.4 – 16.8 ppm (Piedrahita et al., 2014) have been observed for this CO_2 sensor. While studies using this CO sensor have observed RMSEs of 0.10 ppm (Casey et al., 2018) and standard errors 0.28 – 0.44 ppm (Piedrahita et al., 2014). The statistics for each individual sensor and plots of the data are available in the Supplemental Materials (Tables S4 & S5; Figures S6 & S7).

3.2 Variability and Trends Captured by Sensors

Overall, the low-cost sensor systems reflect show similar temporal trends throughout the network, suggesting the presence of community-scale, or possibly regional patterns. Averaging both the CH_4 estimates and the TNMHCs estimates by hour of the day reveals elevated pollutant levels at night and lower levels during the day (Figure S8) – an expected diurnal pattern driven by fluctuations of the planetary boundary layer. The available reference CH_4 data, from site C1 and also plotted in Figure S8, reflects the same diurnal pattern as the sensor systems.

When the regional and local trends are separated using the baseline estimation technique described in Section 2.3.3 the agreement among the network sensors increases for the baseline datasets. This is illustrated in Table 6, which compares the average correlation coefficient for the complete dataset verses the baseline dataset between all sensor system pairs, sensor system pairs deployed in the same neighborhood, and sensor system pairs deployed in relatively close proximity (~ 140 m). Two trends emerge from these results: the first being that the correlation coefficients between baseline datasets are higher than for the complete datasets and the second being that the correlations tend to be higher for sensor systems that are closer together. The complete datasets include the short-term changes in pollutant levels that may be driven by local emissions or other phenomena, which are more likely to differ between sites than the diurnal/regional trends. Additionally, sites closer together are likely to have more shared influences in addition to being impacted more similarly by factors such as meteorology.

Given high level of agreement displayed by these sensor systems when they are co-located or amongst their estimated baseline signals, it stands to reason that when the pollutant levels estimated by these sensors vary the cause may be due to due to exposure to different pollutants, different pollutant mixtures, or differences in environmental factors (i.e., temperature of humidity). For example, a comparison of data collected during the co-location and during the field deployment, shown in Figure 4, reveals the variability that sensors can reflect when placed at different field sites approximately 140 m apart. Despite this close proximity, there are significant increases, well above the expected uncertainties of roughly 0.2 ppm and 50 ppb for CH_4 and TNMHCs respectively, that occur in both the correlated and uncorrelated data from these two field sites (Figure 4, in red). This data suggests that the sensors may be detecting phenomena that affect both sites simultaneously, that affect one site or the other, and that affect the two sites disproportionately. As this data

results from models trained to estimate CH_4 and TNMHCs levels, these increases may be driven by exposure to these target pollutants, though we cannot rule out the influence of confounding pollutants or unexpected effects of environmental conditions. In either case, these sensors appear to be well-suited to identify unique trends that may be the result of small-scale spatial and temporal variability in pollutant concentrations.

3.3. Using Sensor Data to Characterize Potential Sources

While there remains a continual need to better understand the capabilities and limitations of low-cost sensors as well as how to improve data quality, leveraging multi-sensor systems and the comparability of sensors may be able to provide useful preliminary information. Using these networks, we can examine individual events potentially resulting from local emissions, we can study the spatial and temporal patterns across a network, or we can leverage different types of information to explore what a sensor network might suggest about the air quality throughout a community. The following analysis demonstrates these different approaches.

3.3.1 Examining Individual Air Quality Events—Short-term increases in pollutants or episodic events captured by low-cost sensors may be reflective of emissions from local sources. Here several apparent events have been selected to serve as examples of how we may be able to increase our understanding of these possible events. The periods shown in Figures 5 – 7 were selected as they depict increases in the estimated pollutant levels well above the expected uncertainties in addition to differing trends between two field sites that are relatively close together. These examples are not assumed to be representative of the datasets.

The first event discussed is shown in Figure 5, which depicts baseline-removed data from sites E1 and E2, the two sites defined nearest the point source of interest identified by the community. This event is more well-defined in the two multi-sensor models, with changes in the estimated concentrations above the expected uncertainty of ~0.2 ppm for CH_4 and ~50 ppb for TNMHCs. The difference in predicted levels between the single and multi-sensor models for CH_4, as well as the levels predicted by the TNMHC model, may provide insight into the composition of the emissions. Previous research has demonstrated that the Figaro TGS 2602 appears to be responsive to heavier hydrocarbons, such as BTEX compounds, while the Figaro TGS 2600 is responsive to a range VOCs including lighter hydrocarbons, such as CH_4 (Collier-Oxandale et al., 2019). Though it is important to also consider the potential for known or unknown interferents to elicit a response from the sensors as well as environmental conditions unaccounted for by the calibration models.

Multi-sensor systems may provide greater insight into these responses by VOC sensors. In Figure 6, panel a, CO_2 and CO sensor data, from Site E2, have been added to the data shown in Figure 5. Both CO and CO_2 are combustion by-products and the lack of any significant or correlated response between the TNMHCs and CO_2 or CO suggests that the sensors' response to TNMHCs is more likely the result of volatilized or vented hydrocarbons as opposed to hydrocarbons from either complete or incomplete combustion. Additionally, these trends predicted by the VOC sensors are not driven by any obvious environmental

conditions as they do not correlate with the data from the temperature or humidity signals (Figure S9).

Meteorological data may also provide useful context to these measurements. The increases or possible episodic events at Sites E1 and E2 are not correlated and as indicated by the pollution roses in Figure 6, panels b and c, the increases do not appear to share a single origin. The increases at Site E1 appear to originate from the north/northeast and from the west/northwest for Site E2. As volatilized or evaporative hydrocarbons are emissions that may be associated with oil and gas activity (Moore et al., 2014; Warneke et al., 2014) and Site E1 is located to the west of the extraction site (<50 m) while Site E2 is to the east of the extraction site (<50 m), this analysis seems to suggest that the petroleum extraction site is one possible source of this emission event. As previously mentioned, there are no gas stations near these two sensor sites, however other potential sources of volatilized VOCs include a dry cleaner to the south of Site E1 (<100 m) and an auto repair shop to the southwest of E1 (<150 m). Further research using conventional measurement techniques would certainly be necessary to confirm the source of these emissions, but low-cost sensors may provide useful preliminary or indicative information in areas lacking current or historic data.

Two additional examples are shown in Figure 7. During both periods, significant increases in TNMHC estimates occurred, independent of any response in the CO_2 data. Similar to the previous example, this difference in responses may suggest that volatilized hydrocarbons as opposed to those originating from combustion sources are responsible. Additional context is provided by the pollution roses, though in the case of the second example, in panels c and d, the likely origin of the elevations in the TNMHC estimations is less clear.

Figure 8 depicts periods of baseline-removed TNMHC and CO_2 data from sites directly adjacent to a major roadway. In all three panels, early morning, weekday hours are presented and increases in TNMHCs above the excepted uncertainty are observed. During the early morning hours, the planetary boundary layer is still low and facilitates the accumulation of pollutants; this condition combined with morning rush hour traffic may account for some of these increases. Furthermore, the TNMHC trends are well correlated with their respective CO_2 sensor trends as shown by the R^2 values in Table 7. This agreement between the sensor signals indicates that a combustion source or sources are likely responsible for these increases or potential episodic events.

In Figure 8 there are also several short-term increases in the TNMHC estimates that are not correlated with the CO_2 estimates indicating the possible presence of volatilized TNMHCs. Though these instances also reiterate the importance of considering additional information and datasets, like potential nearby sources and meteorological data as these responses may be the result of exposure to the target pollutants, known or unknown interferents, or a sudden change in environmental conditions. Considering the larger spatial and temporal patterns across a network can provide greater insight into responses by the sensor systems.

3.3.2 Identifying Temporal and Spatial Patterns Related to Local Air Quality Events—Examining patterns associated with short-term increases in the sensor responses

may provide insights into where and when localized exposure events occur. In Figure 9 the complete data from four sites is sorted by hour of the day. Site R3 is above a busy roadway, Site N5 serves as a neighborhood site as it is in a predominantly residential area, approximately 250 m from the nearest highway, and approximately 950 m from the nearest petroleum extraction site. This site is also less than 100 m from an autobody shop and less than 400 m from several gas stations. The final two sites shown, E1 and E2, are again at the location of interest identified by the community. Similar box plots are available for each field site in the Supplemental Materials (Figures. S10 – S12).

Across all four sites shown in Figure 9, elevated levels of CH_4 and TNMHCs occur during the early morning hours. This trend is likely driven by a combination of morning rush hour traffic and the lowered planetary boundary layer. Elevated concentrations during the night are also common to the sites; these are likely driven primarily by the lowering of the planetary boundary layer at night reducing dispersion. Regarding the early morning increases observed in the CH_4 estimates, this response is likely driven by interferents. Previous work has demonstrated that, in addition to various VOCs, the Figaro TGS 2600 is responsive to CO (Collier-Oxandale et al., 2018a & 2019); a pollutant that is likely to be associated with emissions from vehicles.

In addition to these expected diurnal patterns, there are also increases in the late-morning and afternoon, above the expected uncertainty for each pollutant, that are not consistent with expected traffic patterns and likely the result of a different source. In Figure 9, these increases or possible episodic events occur most prominently at site E2 and to a lesser extent at site E1. Similar trends are observed at other sites in the network as well, including N1 and N4 (Figures S10 – S12). For the four sites shown here, these repeated increases even result in elevated 95^{th} percentile values in the afternoon hours for Site E2 (Figure S13). Though the highest overall 95^{th} percentile value for TNMHCs occurs at site N4 (62.5 ppb), followed by site E2 (61.0 ppb), and then site N1 (59.7 ppb).

3.3.3 Leveraging Sensor Networks to Learn More about Potential Sources in a Community—Together these patterns of how and when responses are observed from multi-sensor devices provide useful information about local air quality patterns. Figure 11 provides information for the whole network, depicting the baseline-removed CO_2 and TNMHC data, colored by the hour of the day, for each sensor system. Similar plots are available for CO and TNMHCs and CO and CO_2 (Figures S14 – S15). A reference ratio has also been added to the second set of plots that represent the expected CO:CO_2 ratio based on the CARB emission inventory for the South California Air Basin as observed during aircraft measurements (Peischl et al., 2013). This additional information illustrates how the observation drawn from the CO and CO_2 sensors appear to be in agreement with expected trends and previous studies.

The patterns in Figure 11 again suggest that the sensors may be responding to hydrocarbons from both vented or volatilized sources and combustion sources. At sites N1, N4, and E2 there are relatively large increases in hydrocarbons, reaching levels greater than five times the expected uncertainty, and independent of any apparent increase in CO_2. As these plots depict the baseline-removed data, these increases are likely to be short-term events as

opposed to the result of regional or diurnal pollutant trends. Whether these sensors are responding to the target pollutants, interferents, or a mixture, the dramatic differences in patterns between these three sites and the other sites may warrant further investigation with more conventional sampling techniques. There are also interesting differences amongst sites where TNMHCs and CO_2 seem to be positively related. For example, at some sites the two pollutants are correlated along a more consistent ratio (e.g., N2 or R2), while at other sites there appear to be more varied relationships between the two pollutants (e.g., N5 or R5). One explanation for this difference could be the presence of more similar verses more varied combustion sources. Site R5 was located at a garage where maintenance vehicles were kept, and a wider array of gasoline and heavy-duty vehicles entering and leaving the facility might account for the larger variance in the TNMHC:CO_2 ratio.

In Figure 12, the concentration estimates for CH_4 and CO are included to provide further insight into these increases or possible events. Similar to Figure 11, there are increases in estimated CH_4 during the afternoon and early evening hours at site E2 that occur independent of any apparent increase in CO. Alternatively, at site R4 there do not appear to be any increases in CH_4 that are independent of increases in CO. This example confirms that while one of the VOC sensors used is cross-sensitive to CO the presence of CO does not account for at least some of the observed increases at site E2. This also further supports the assertion that one or more local non-combustion source (i.e., not from traffic or vehicle exhaust) may be contributing to elevated hydrocarbon levels at this site. Though the calibration model is trained to estimate CH_4, these non-combustion elevations may be driven by other VOCs, a mixture, or possibly other interferents. In terms of the varied and distinct ratios observed between the CH_4 and TNMHC estimates, these may indicate the presence of different VOCs or different VOC mixtures and certain ratios may even be associated with specific sources (Collier-Oxandale, et al., 2019). The ability to measure CH_4 may be especially valuable for community-based air quality monitoring as this compound may be able to serve as a marker for certain source types, such as sources associated with oil and gas development.

Figure 13 summarizes the spatial variability demonstrated by the previous analysis with each map depicting the average TNMHC:CO_2 ratio for a given time of day (i.e., the morning hours, afternoon hours, and evening hours). The ratios are lower and more similar in the morning hours, but exhibit more variability during the afternoon and evening hours. At sites R2 and R4, directly adjacent to a major roadway, the ratios do not increase that dramatically. By comparison there are several neighborhood sites, such as E2 (circled in yellow in Figure 13), where the short-term increases observed in the baseline-removed data result in higher averages. Using this type of approach to leverage the information available from a dense network of multi-sensor devices could be a way to characterize an area and provide preliminary information regarding potential sources or anomalies in local trends.

3.4 Leveraging Sensor Data to Understand Resident Experiences

Given the accessibility of low-cost sensor systems, it is likely that communities concerned about their local air quality will be interested in using this technology to better understand their local environment. If sensor systems are leveraged appropriately, with communication

of sensor limitations and careful interpretation of sensor data, this may be accomplished. For example, sensor networks may be able to help identify the locations of “hotspots” or recurring episodic events, which could then be compared to residents’ observations. In this type of application, sensor data may be able to provide compelling evidence that sources of concern are worthy of further investigations. It is also possible that a sensor may not be able to detect a particular source or that a sensor may reflect relatively low emissions from a source. In all of these cases, sensor data has the potential to provide new information to facilitate further investigation, a new approach to that investigation, or informed action with the aim of improving community health.

Having examined the low-cost sensor data, we wished to consider how this data related to observations from community residents. As described previously, our community partners provided observational information from residents living near the petroleum extraction site. To begin, the sensor data was examined for potential air quality events corresponding to that observational data; several examples are shown in Figure 14. Many of these examples include increases in TNMHC estimates that are independent of significant changes in CO_2, possibly indicating a volatilized or vented source

While this is not conclusive, as there are several peaks in the measured concentration that do not coincide with residents’ observations, this dataset provides a unique opportunity to pilot ways to access and compare the information available in these two distinct datasets. One approach is demonstrated in Tables 8 and 9, where the data has been summarized using a confusion matrix. In this matrix, this study has been sorted into days where both an observation and an increase in estimated pollutant levels occurred, days where one or the other occurred, or days where neither occurred. In this example analysis a “peak” is defined as increase in estimated TNMHCs greater than 100 ppb in the baseline-removed data, and given the frequency of the peaks, days on which more than three occurred were used for this comparison. Table 8 compares the resident observations with data from the two nearby sampling sites and Table 9 compares these observations with data from sites further away.

In general, more peaks appear to have occurred at sites E1 and E2 as compared to sites R3 and R4. Comparing Tables 8 and 9, a higher rate or both true positive results and false positive results was seen when comparing residents’ observations to data from the nearby sites. Though false positive results could be explained by peaks that occur when residents are away from home or asleep. Additionally, a higher rate of false negative results occurred when comparing the residents’ observations to the data from far sites. This initial analysis seems to indicate greater agreement between the residents’ observations and sensor data collected at the nearby sampling sites.

Ways to improve this type of analysis would include conducting a more comprehensive data collection from residents living throughout the area of the sensor network, systematically comparing data from many residents to sensor data throughout the network and testing different ways of defining peaks and responses in the sensor data. Particularly in high density areas with many sources, like Los Angeles, there is the potential for sensor data to support a better understanding of local air quality as well as the potential for observational information to provide additional context to the air quality trends observed and their impacts

on nearby populations. While additional considerations around how to collect and interpret this observational information or local knowledge are necessary, it is possible that leveraging these two types of information could facilitate more participatory and inclusive studies; while also enhancing data interpretation leading to more locally relevant and actionable results for a community.

4. CONCLUSION

In line with previous research utilizing the same VOC sensors, the results of sensor performance quantification were similar to expected results and the converted sensor data reflected expected environmental trends – specifically regional and diurnal trends (Collier-Oxandale et al., 2018a & 2019). This network also provides new information by revealing short-term or episodic enhancements over background in CH_4 and TNMHC concentrations that are likely the result of local emission events given the time of day at which many of them occur. Incorporating the CO_2 and CO sensor data from the same time periods seems to suggest that some of these increases were the result of volatilized or vented hydrocarbons, as opposed to a combustion source. For the location of interest (near sites E1 and E2), some of these increases even coincide with observations made by nearby residents concerning odors or activity at the petroleum extraction facility. While these results are not conclusive, particularly given the limitations of low-cost sensors, they seem to indicate that combustion sources, such as traffic, are not the only source of hydrocarbons having a measurable impact on air quality in this community. These results also demonstrate how multi-pollutant sensor networks may be able to help meet the desire for more air quality data at the neighborhood-level. We used this approach to distinguish between emissions originating from different sources-types within a community, and we hope that others continue to build on this work – leading to more locally relevant information for communities that can support actions to improve public health and reduce disparities in exposure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Funding provided through MetaSense NSF Grant CNS-1446912, NSF CyberSEES (1442971), NSF AWG SRN (CBET-1240584), and NIEHS (R21ES027695). We would also like to thank all project partners at the University of Southern California Keck School of Medicine, Redeemer Community Partnership, William Flores, Esperanza Community Housing, and Occidental College. Additional acknowledgements to all community residents who assisted by hosting a Y-Pod sensor system. We would also like to thank regulatory partners at South Coast Air Quality Management District and San Joaquin Valley Air District/California Air Resources Board for this assistance with co-locations. Thanks to Christine Wiedinmyer (CIRES) for the deployment map included in this paper (Section 2.1). And a final thanks to current and former members of the Hannigan Research Lab, especially Nicholas Masson and Drew Meyers for their work on the U-Pod/Y-Pod hardware and software, and Evan Coffey and Kira Sadighi for their assistance with the deployment in Los Angeles.

REFERENCES

Adgate JL, Goldstein BD, & McKenzie LM (2014): Potential public health hazards, exposures and health effects from unconventional natural gas development, Environ. Sci. Technol, 48(15), 8307–8320, doi:10.1021/es404621d [PubMed: 24564405]

Bamberger I, Stieger J, Buchmann N, & Eugster W (2014). Spatial variability of methane: Attributing atmospheric concentrations to emissions. Environmental Pollution, 190, 65–74. 10.1016/j.envpol.2014.03.028 [PubMed: 24727588]

Beloff BR, Beaver ER, & Massin H (2000). Assessing societal costs associated with environmental impacts. Environmental Quality Management, 10(2), 67–82. 10.1002/1520-6483(200024)10:2<67::AID-TQEM8>3.0.CO;2-B

Brown P (1992). Popular Epidemiology and Toxic Waste Contamination: Lay and Professional Ways of Knowing. Journal of Health and Social Behavior, 33(3), 267–281. [PubMed: 1401851]

Brugge D, Durant JL, & Rioux C (2007). Near-highway pollutants in motor vehicle exhaust: A review of epidemiologic evidence of cardiac and pulmonary health risks. Environmental Health: A Global Access Science Source, 6(23), 1–12. 10.1186/1476-069X-6-23 [PubMed: 17233918]

Casey JG, Ortega J, Coffey E, & Hannigan M (2018). Low-cost measurement techniques to characterize the influence of home heating fuel on carbon monoxide in Navajo homes. Science of the Total Environment, 625, 608–618. 10.1016/j.scitotenv.2017.12.312

Casey JG and Hannigan MP (2018). Testing the performance of field calibration techniques for low-cost gas sensors in new deployment locations: across a county line and across Colorado, Atmos. Meas. Tech, 11, 6351–6378, 10.5194/amt-11-6351-2018.

Casey JG, Collier-Oxandale A, & Hannigan M (2019). Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. Sensors and Actuators B: Chemical, 283, 504–514. 10.1016/J.SNB.2018.12.049

Castell N, Dauge FR, Schneider P, Vogt M, Lerner U, Fishbain B, … Bartonova A (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environment International, 99, 293–302. 10.1016/j.envint.2016.12.007 [PubMed: 28038970]

Cheadle L, Deanes L, Sadighi K, Casey JG, Collier-Oxandale A, & Hannigan M (2017). Quantifying Neighborhood-Scale Spatial Variations of Ozone at Open Space and Urban Sites in Boulder, Colorado Using Low-Cost Sensor Technology. Sensors, 17. 10.3390/s17092072

Chilingar GV, & Endres B (2005). Environmental hazards posed by the Los Angeles Basin urban oilfields: An historical perspective of lessons learned. Environmental Geology, 47(2), 302–317. 10.1007/s00254-004-1159-0

Clark LP, Millet DB, & Marshall JD (2017). Changes in transportation-related air pollution exposures by race-ethnicity and socioeconomic status: Outdoor nitrogen dioxide in the United States in 2000 and 2010. Environmental Health Perspectives. 10.1289/EHP959

Clements AL, Griswold WG, RS A, Johnston JE, Herting MM, Thorson J, … Hannigan M (2017). Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary). Sensors, 17. 10.3390/s17112478

Collier-Oxandale A, Hannigan MP, Casey JG, Piedrahita R, Halliday HS, & Johnston J (2018a). Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. Atmospheric Measurement Techniques, 11, 3569–3594. 10.5194/amt-2017-421 [PubMed: 33442426]

Collier-Oxandale A, Coffey E, Thorson J, Johnston J, & Hannigan M (2018b). Comparing Building and Neighborhood-Scale Variability of CO2 and O3 to Inform Deployment Considerations for Low-Cost Sensor System Use. Sensors, 18(5), 1349. 10.3390/s18051349

Collier-Oxandale AM, Thorson J, Halliday H, Milford J, and Hannigan M (2109). Understanding the ability of low-cost MOx sensors to quantify ambient VOCs, Atmos. Meas. Tech, 12, 1441–1460, 10.5194/amt-12-1441-2019.

Cross ES, Williams LR, Lewis DK, Magoon GR, Onasch TB, Kaminsky ML, … Jayne JT (2017). Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements, 3575–3588.

De Vito S, Massera E, Piga M, Martinotto L, & Di Francia G (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors and Actuators, B: Chemical, 129, 750–757. 10.1016/j.snb.2007.09.060

De Vito S, Fattoruso G, Liguoro R, Oliviero A, Massera E, Sansone C, … Di Francia G (2011). Cooperative 3D Air Quality assessment with wireless chemical sensing networks. Procedia Engineering.

Eapoaito E, De Vito S, Salvato M, Fattoruso G, Castell N, Karatzas K, & Di Francia G (2017, 5). Is on field calibration strategy robust to relocation?. In 2017 ISOCS/IEEE International Symposium on Olfaction and Electronic Nose (ISOEN) (pp. 1–3). IEEE.

English PB, Olmedo L, Bejarano E, Lugo H, Murillo E, Seto E, … Northcross A (2016). The Imperial County Community Air Monitoring Network : A Model for Community-based Environmental Monitoring for Public Health Action, 1–5.

Etzion Y, & Broday D (2018). Highly resolved spatiotemporal variability of fine particle number concentrations in an urban neighborhood. Journal of Aerosol Science.

Eugster W, & Kling GW (2012). Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies. Atmospheric Measurement Techniques, 5(8), 1925–1934. 10.5194/amt-5-1925-2012

Figaro, Inc. 2015a. TGS 2602 Data Sheet.

Figaro, Inc. 2015b. TGS 2600 Datasheet.

Fischer PH, Hoek G, Van Reeuwijk H, Briggs DJ, Lebret E, Van Wijnen JH, … Elliott PE (2000). Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam. Atmospheric Environment, 34(22), 3713–3722. 10.1016/S1352-2310(00)00067-4

Hagan DH, Isaacman-vanwertz G, Franklin JP, Wallace LMM, Kocar BD, Heald CL, & Kroll JH (2018). Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments, 315–328

Heimann I, Bright VB, McLeod MW, Mead MI, Popoola OAM, Stewart GB, & Jones RL (2015). Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. Atmospheric Environment, 113, 10–19. 10.1016/j.atmosenv.2015.04.057

Helmig D, Thompson CR, Evans J, Boylan P, Hueber J, and Park JH: Highly elevated atmospheric levels of volatile organic compounds in the uintah basin, Utah, Environ. Sci. Technol, 48, 4707–4715, doi:10.1021/es405046r, 2014. [PubMed: 24624890]

Jerrett M, Donaire-gonzalez D, Popoola O, Jones R, Cohen RC, Almanza E, … Nieuwenhuijsen M (2017). Validating novel air pollution sensors to improve exposure estimates for epidemiological analyses and citizen science. Environmental Research, 158, 286–294. 10.1016/j.envres.2017.04.023 [PubMed: 28667855]

Kizel F, Etzion Y, Shafran-Nathan R, Levy I, Fishbain B, Bartonova A, & Broday D (2018). Node-to-node field calibration of wireless distributed air pollution sensor network. Environmental Pollution.

Künzli N, McConnell R, Bates D, Bastain T, Hricko A, Lurmann F, … & Peters J (2003). Breathless in Los Angeles: the exhausting search for clean air. American Journal of Public Health, 93(9), 1494–1499. [PubMed: 12948969]

Kwon J, Weisel CP, Turpin BJ, Zhang J, Korn LR, Morandi MT, … Colome S (2006). Source Proximity and Outdoor-Residential VOC Concentrations: Results from the RIOPA Study. Environmental Science and Technology, 40(13), 4074–4082. 10.1021/es051828u [PubMed: 16856719]

Leidinger M, Sauerwald T, Conrad T, Reimringer W, Ventura G, & Schütze A (2014). Selective Detection of Hazardous Indoor VOCs Using Metal Oxide Gas Sensors. Procedia Engineering, 87, 1449–1452. 10.1016/j.proeng.2014.11.722

Lewis AC, Lee J, Edwards PM, Shaw MD, Evans MJ, Moller SJ, … Buckley JW (2016). Evaluating the performance of low cost chemical sensors for air pollution research. The Royal Society of Chemistry, Faraday Di. 10.1039/C5FD00201J

Loh MM, Levy JI, Spengler JD, Houseman EA, & Bennett DH (2007). Ranking cancer risks of organic hazardous air pollutants in the United States. Environmental Health Perspectives, 115(8), 1160–1168. 10.1289/ehp.9884 [PubMed: 17687442]

Marshall JD (2008). Environmental inequality: Air pollution exposures in California's South Coast Air Basin. Atmospheric Environment, 42, 5499–5503. 10.1016/j.atmosenv.2008.02.005

Masson N, Piedrahita R, & Hannigan M (2015a). Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring. Sensors and Actuators B: Chemical, 208, 339–345. 10.1016/j.snb.2014.11.032

Masson N, Piedrahita R, & Hannigan M (2015b). Quantification method for electrolytic sensors in long-term monitoring of ambient air quality, Sensors 1–17.

Mead MI, Popoola O. a. M., Stewart GB, Landshoff P, Calleja M, Hayes M, … Jones RL (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmospheric Environment, 70(2), 186–203. 10.1016/j.atmosenv.2012.11.060

Miranda ML, Edwards SE, Keating MH, & Paul CJ (2011). Making the environmental justice grade: the relative burden of air pollution exposure in the United States. International journal of environmental research and public health, 8(6), 1755–1771. doi:10.3390/ijerph8061755 [PubMed: 21776200]

Moltchanov S, Levy I, Etzion Y, Lerner U, Broday D, & Fishbain B (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. Science of the Total Environment.

Moore CW, Zielinska B, Petron G, & Jackson RB: Air impacts of increased natural gas acquisition, processing, and use: a critical review. Environ. Sci. Technol, 48(15), 8349–8359, doi:10.1021/es4053472, 2014. [PubMed: 24588259]

Mukerjee S, Smith LA, Thoma ED, Oliver KD, Whitaker DA, Wu T, … Stallings C (2016). Spatial analysis of volatile organic compounds in South Philadelphia using passive samplers. Journal of the Air & Waste Management Association, 66(5), 492–498. 10.1080/10962247.2016.1147505 [PubMed: 26828464]

Peischl J, Ryerson TB, Brioude J, Aikin KC, Andrews AE, Atlas E, … Parrish DD (2013). Quantifying sources of methane using light alkanes in the Los Angeles basin, California. Journal of Geophysical Research Atmospheres, 118, 4974–4990. 10.1002/jgrd.50413

Piedrahita R, Xiang Y, Masson N, Ortega J, Collier A, Jiang Y, … Shang L (2014). The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmospheric Measurement Techniques Discussions, 7(2), 2425–2457. 10.5194/amtd-7-2425-2014

Ruckstuhl AF, Henne S, Reimann S, Steinbacher M, Vollmer MK, Doherty SO, & Buchmann B (2012). Robust extraction of baseline signal of atmospheric trace species using local regression, 2613–2624. 10.5194/amt-5-2613-2012

Sadd J, Shamasunder B: Oil extraction in Los Angeles: health, land use, and environmental justice consequences, from the Drilling Down Report, by the Liberty Hill Foundation, report available at https://www.libertyhill.org/sites/libertyhillfoundation/files/Drilling%20Down%20Report_1.pdf, 2015.

Sadighi K, Coffey E, Polidori A, Feenstra B, Lv Q, & Henze DK (2018). Intra-urban spatial variability of surface ozone in Riverside, CA : viability and validation of low-cost sensors, 1777–1792.

Sauerwald T, Baur T, Leidinger M, Reimringer W, Spinelle L, Gerboles M, … Schütze A (2018). Highly sensitive benzene detection with metal oxide semiconductor gas sensors – An inter-laboratory comparison. Journal of Sensors and Sensor Systems, 7(1), 235–243. 10.5194/jsss-7-235-2018

Shamasunder B, Collier-oxandale A, Blickley J, Sadd J, Chan M, Navarro S, … Wong NJ (2018). Community-Based Health and Exposure Study around Urban Oil Developments in South Los Angeles. International

Smith K, Edwards P, Evans M, Lee J, Shaw M, Squires F, … Lewis A (2017). Clustering approaches to improve the performance of low cost air pollution sensors. Faraday Discussions, 200, 621–637. Royal Society of Chemistry.Journal of Environmental Research and Public Health, 15(138). 10.3390/ijerph15010138 [PubMed: 28608899]

Snyder EG, Watkins TH, Solomon PA, Thoma ED, Williams RW, Hagler GSW, … Preuss PW (2013). The changing paradigm of air pollution monitoring. Environmental Science & Technology, 47(20), 11369–77. 10.1021/es4022602 [PubMed: 23980922]

Souza JCD, Jia C, Mukherjee B, & Batterman S (2009). Ethnicity, housing and personal factors as determinants of VOC exposures. Atmospheric Environment, 43(18), 2884–2892. 10.1016/j.atmosenv.2009.03.017

Spinelle L, Gerboles M, Kok G, Persijn S, & Sauerwald T (2017). Review of Portable and Low-Cost Sensors for the Volatile Organic Compounds. Sensors, 17. 10.3390/s17071520

Sun Y-F, Liu S-B, Meng F-L, Liu J-Y, Jin Z, Kong L-T, & Liu J-H (2012). Metal Oxide Nanostructures and Their Gas Sensing Properties: A Review. Sensors, 12, 2610–2631. 10.3390/s120302610 [PubMed: 22736968]

Thoma ED, Brantley HL, Oliver KD, Whitaker DA, Mukerjee S, Mitchell B, … Philadelphia S (2016). South Philadelphia passive sampler and sensor study. Journal of the Air & Waste Management Association, 66(10), 959–970. 10.1080/10962247.2016.1184724 [PubMed: 27192142]

Thorson J, Collier-Oxandale A, & Hannigan M (2019). Using A Low-Cost Sensor Array and Machine Learning Techniques to Detect Complex Pollutant Mixtures and Identify Likely Sources. Sensors, 19(17), 3723.

Vikram S, Collier-Oxandale A, Ostertag M, Menarini M, Chermak C, Dasgupta S, … Griswold W (2019, 2 12). Evaluating and Improving the Reliability of Gas-Phase Sensor System Calibrations Across New Locations for Ambient Measurements and Personal Exposure Monitoring. Atmospheric Measurement Techniques Discussions, 1–38.

Wang C, Yin L, Zhang L, Xiang D, & Gao R (2010). Metal Oxide Gas Sensors: Sensitivity and Influencing Factors. Sensors, 10, 2088–2106. 10.3390/s100302088 [PubMed: 22294916]

Warneke C, Geiger F, Edwards PM, Dube W, Pétron G, Kofler J, … Brown SS (2014). Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah : oil and gas well pad emissions compared to ambient air composition, (x), 10977–10988. 10.5194/acp-14-10977-2014

Yuval et al., Application of a sensor network of low cost optical particle counters for assessing the impact of quarry emissions on its vicinity. Atmospheric Environment, 211:29–17, 2019.

Zimmerman N, Presto AA, Kumar SPN, Gu J, Hauryliuk A, Robinson ES, … Subramanian R (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, 291–313.

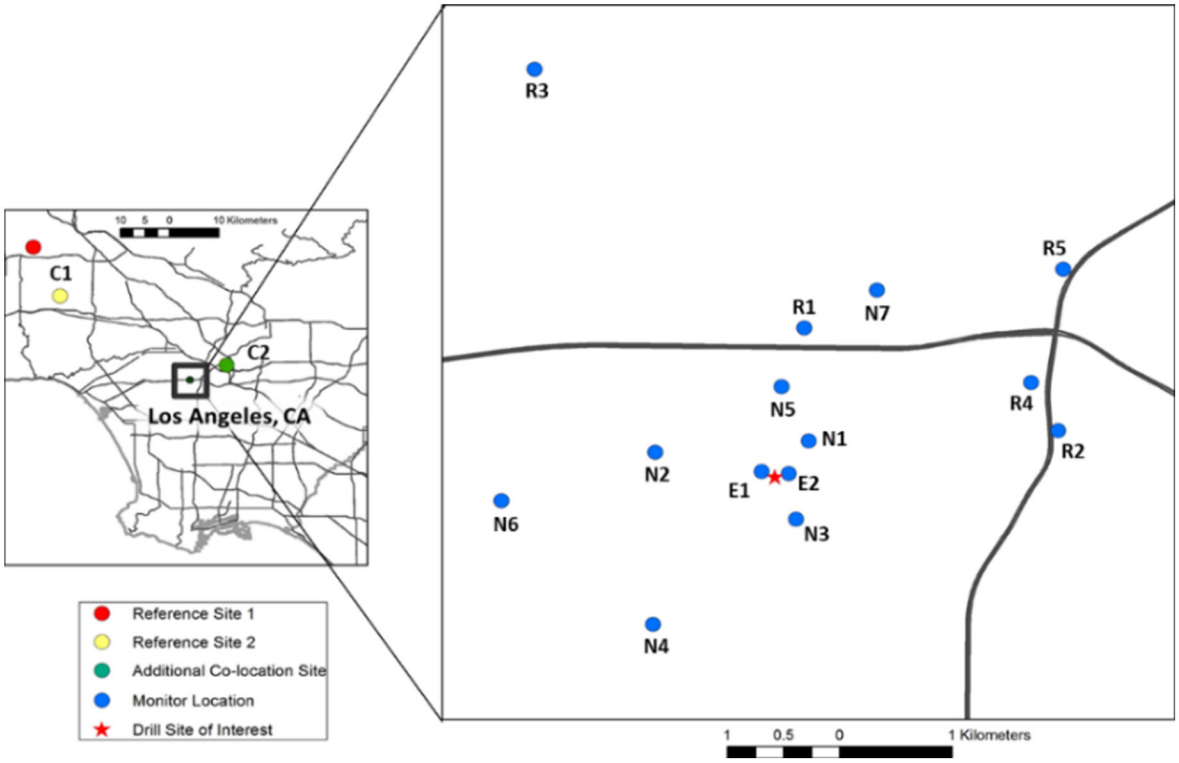

Figure 1:

Map of sampling sites and major sources of interest. Note that the Y-Pod locations have been approximated to the center of their respective blocks in order to protect participant identities. The first initial is indicative of the site type: N – neighborhood site, R – near roadway site, E – near oil extraction site, C – ongoing co-location site (note, Reference Site 1 was not used throughout the field deployment and thus does not have an additional label associated with a Y-Pod).

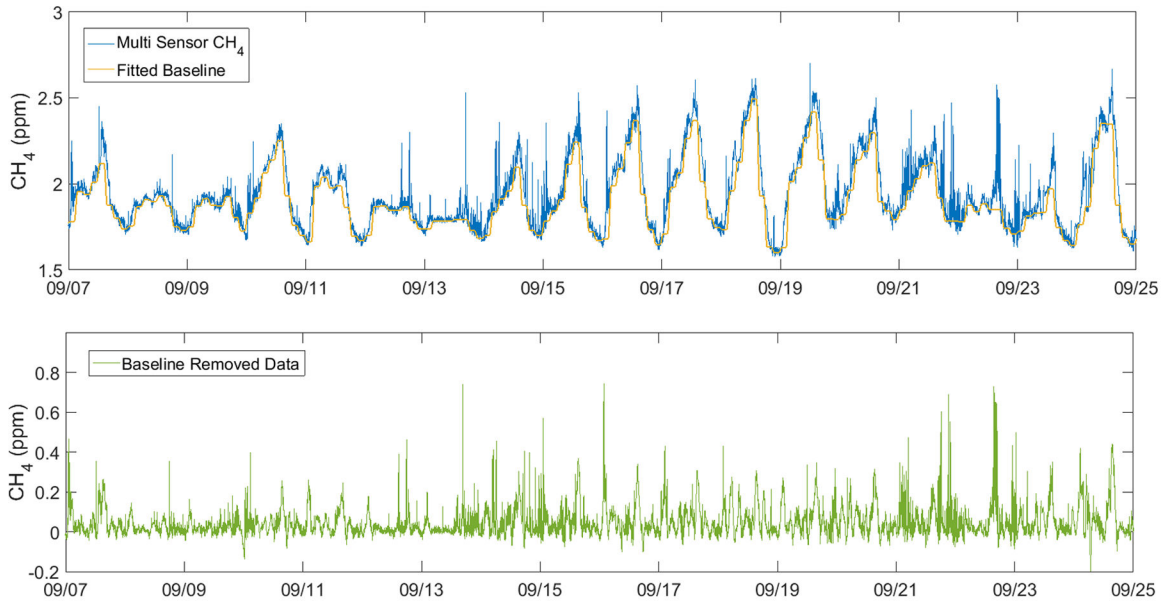

Figure 2:
Illustration of the baseline identification and removal approach

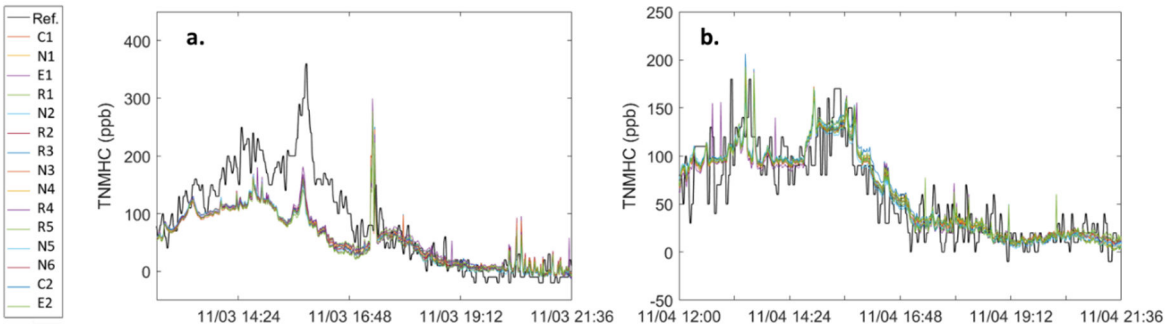

Figure 3:

Excerpts from the co-location period depicting reference data and fitted sensor data for TNMHCs.

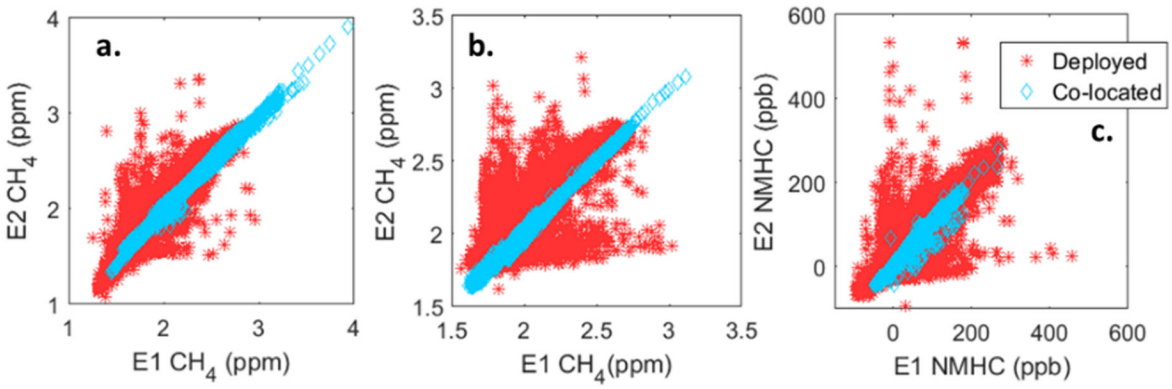

Figure 4:

Paired data for the two sampling sites near the drill site (E1 and E2), with co-located data plotted in blue and deployed data in red; using the a) single sensor CH_4, b) multi-sensor CH_4, and c) multi-sensor NMHC calibration models

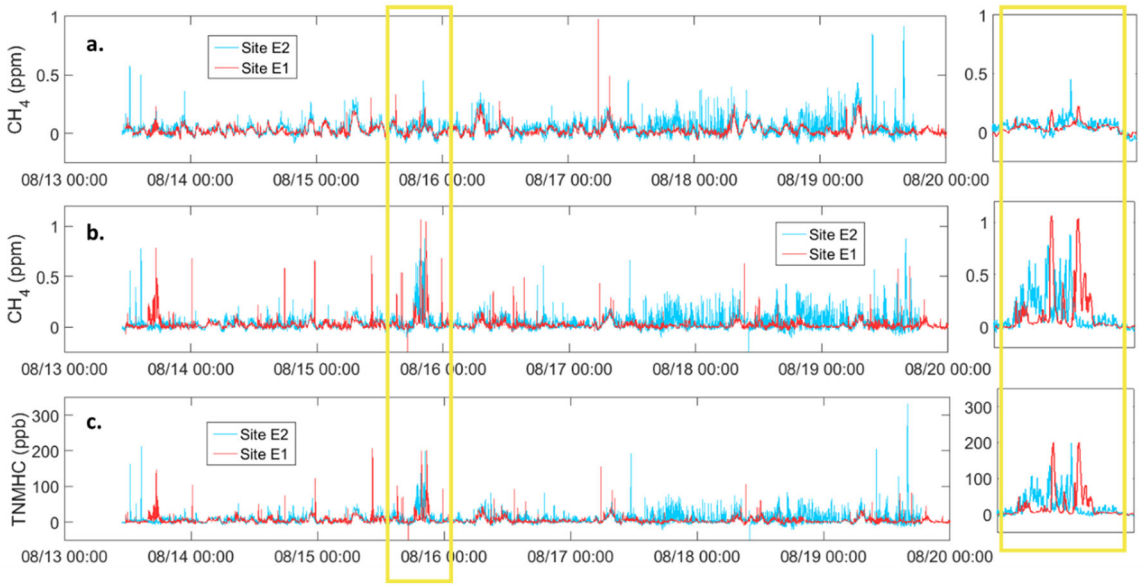

Figure 5:

Approximately one week of baseline-removed data, with panel a) including the single sensor CH_4 data, panel b) including the multi-sensor CH_4 data, and panel c) including the multi-sensor NMCH data. The time stamp is local, and the yellow box highlights the event discussed in the text. To the right of each panel is a zoomed in version of the event highlighted in yellow.

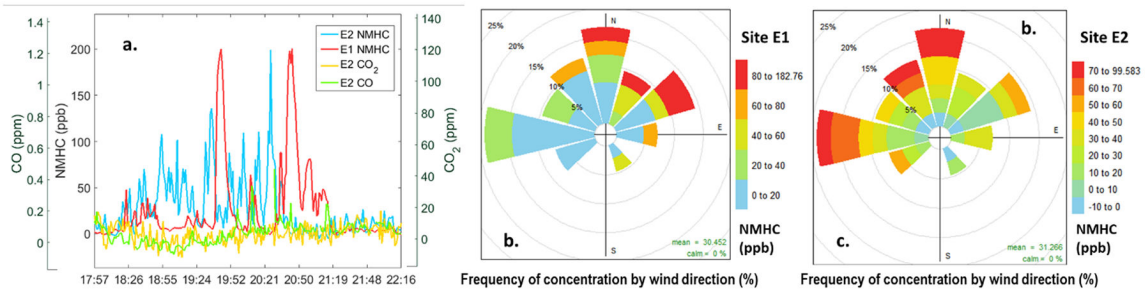

Figure 6:

Panel a) includes baseline-removed TNMHC, CO_2, and CO data from Site E2 as well as baseline-removed TNMHC data from Site E1. The data is from 8/15/16 and the times listed are local times. Panels b) and c) include pollution roses for the period of enhancements in TNMHCs for Sites E1 and E2 respectively, the data used to generate the pollution roses is from precisely the same time period depicted in panel a.

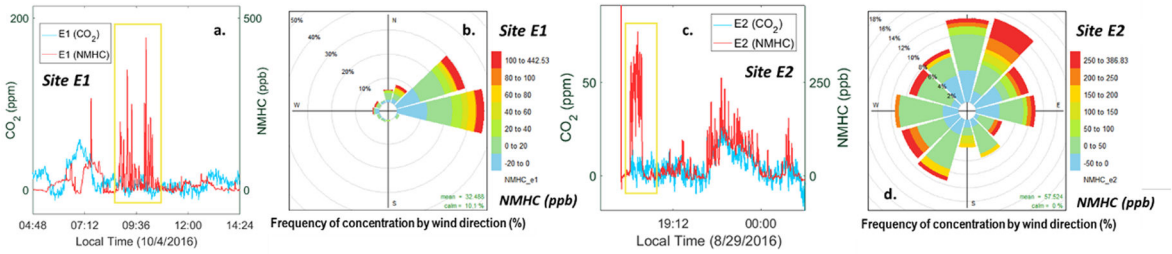

Figure 7:

Panels a) and c) include baseline-removed TNMHC and CO_2 data from Sites E1 and E2 respectively. Panels b) and d) depict pollution roses for the periods in panels a) and c) respectively.

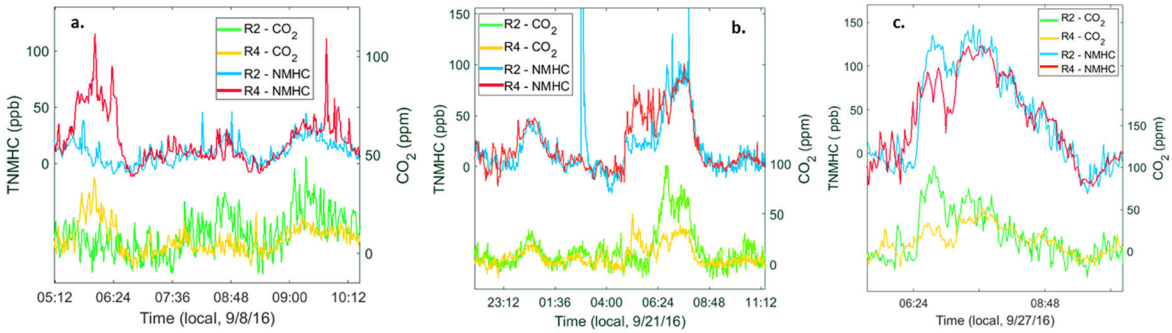

Figure 8:

Panels a – c) include baseline-removed TNMHC and CO_2 data from Sites R2 and R4 from three weekday mornings, during the deployment.

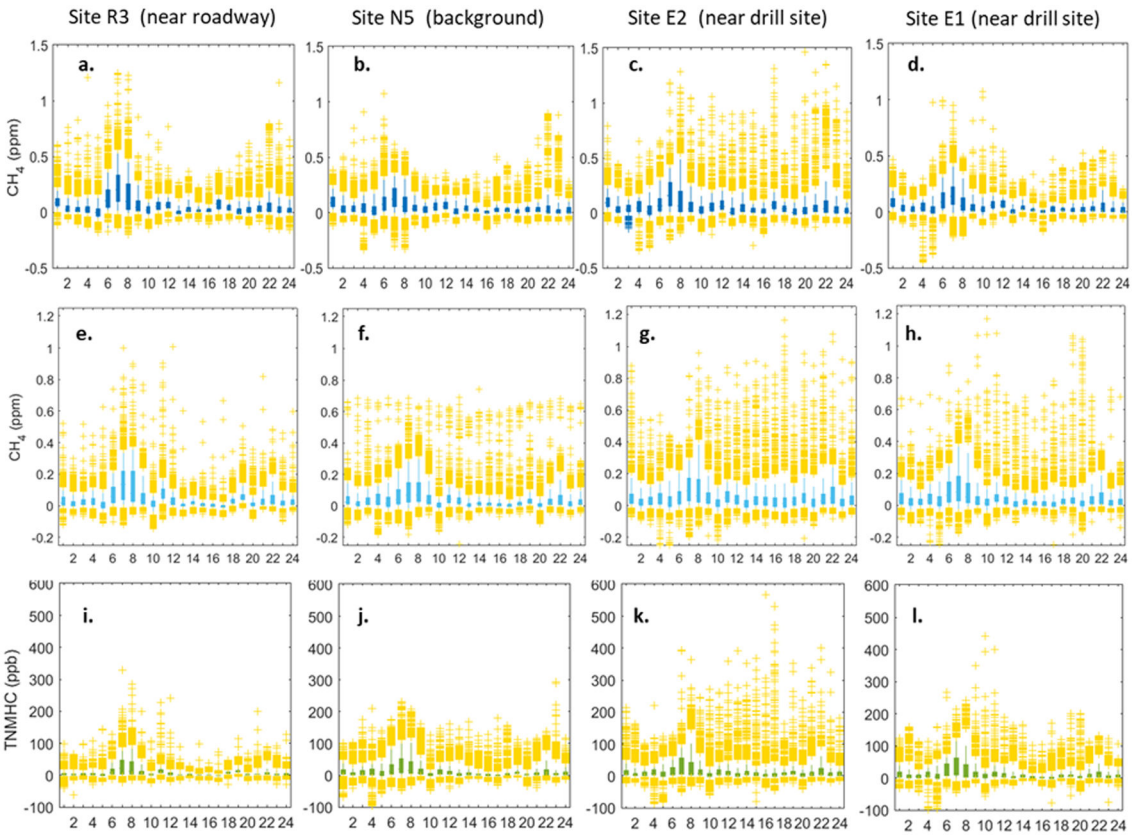

Figure 9. Baseline-removed data from four sites grouped by hour of the day (in local time), panels a-d) include the single sensor CH_4 data, panels e-h include the multi-sensor CH_4 data, and panels i-l) include the multi-sensor NMHC data. The whiskers on the box plots represent the 5^{th} and 95^{th} percentiles respectively. The top and bottom 5^{th} percent of the data are indicated by yellow markers across all plots.

Referencing data from other types of sensors, similar to the examples in Section 3.3.1, again reveals periods of agreement and disagreement between the sensor responses. The increases in the early morning and evening hours seen at the four sites in Figure 9 also appear in the CO_2 and CO sensor data shown in Figure 10. However, the increases in the late-morning and afternoon hours are not reflected in the data for these two pollutants. The timing of this pattern may provide insight into potential sources. Researchers studying the spatial variability of CH_4 determined that increases in ground level CH_4 can be indicative of sources up to 8.4 km away at night, but only 240 m away during the day (Bamberger et al., 2014). This study emphasized the ability of daytime mixing to disperse pollutants, suggesting that increases or episodic events observed during the afternoon are likely the result of relatively local sources as opposed to more distant sources. Thus, it is possible that the afternoon increases observed during this study are not only the results of a vented or volatilized source, but also the result of local emissions.

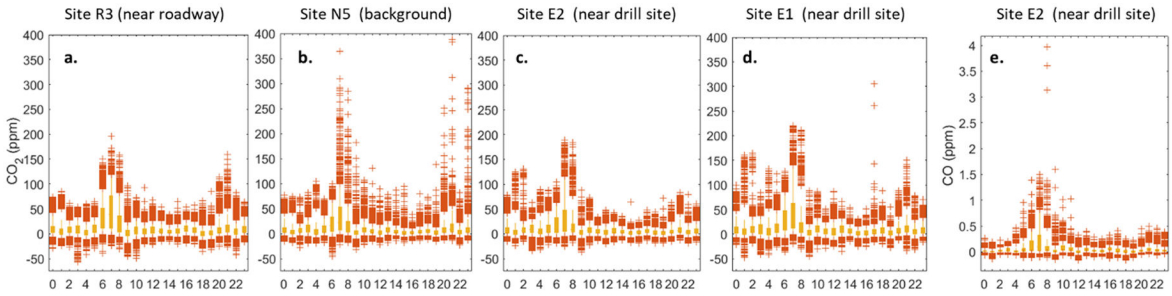

Figure 10:

Baseline-removed data from four sites grouped by hour of the day (in local time), panels a-d include CO_2 data, and panel e includes the CO data collected at site E2.

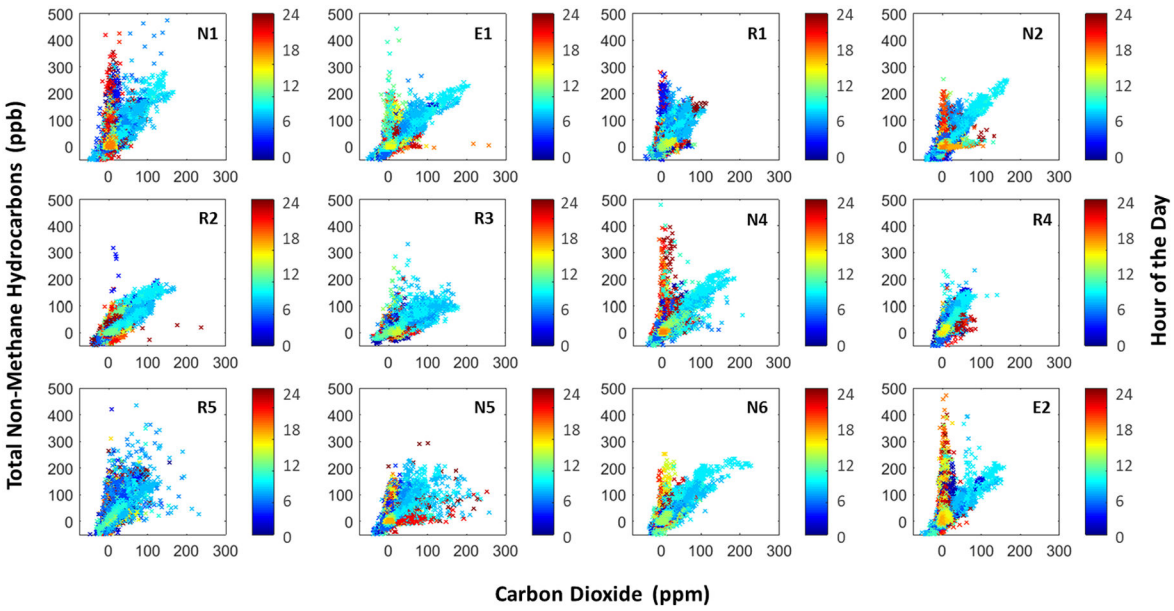

Figure 11:

Scatterplots of baseline-removed data CO_2 and TNMHC data for sites in the deployment area, with the color indicating the hour of the day in local time.

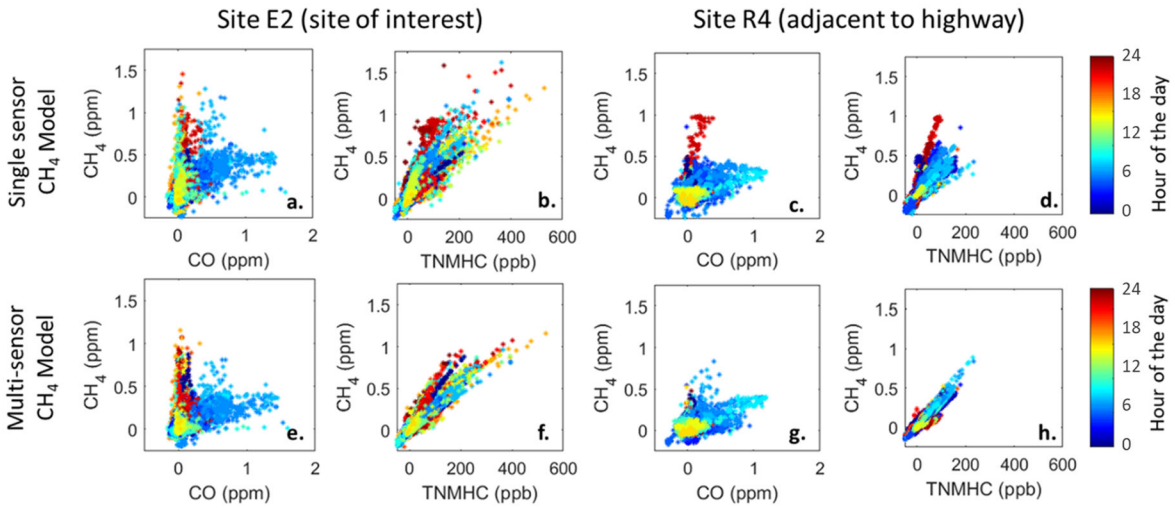

Figure 12:

Scatterplots of baseline-removed data CO or TNMHC data versus CH_4 data from two sites in the deployment area (E2 and R4). Note the CH_4 on the top row was converted using the single-sensor calibration model and the CH_4 values in the bottom row were converted using the multi-sensor calibration model. The color indicates the hour of the day in local time.

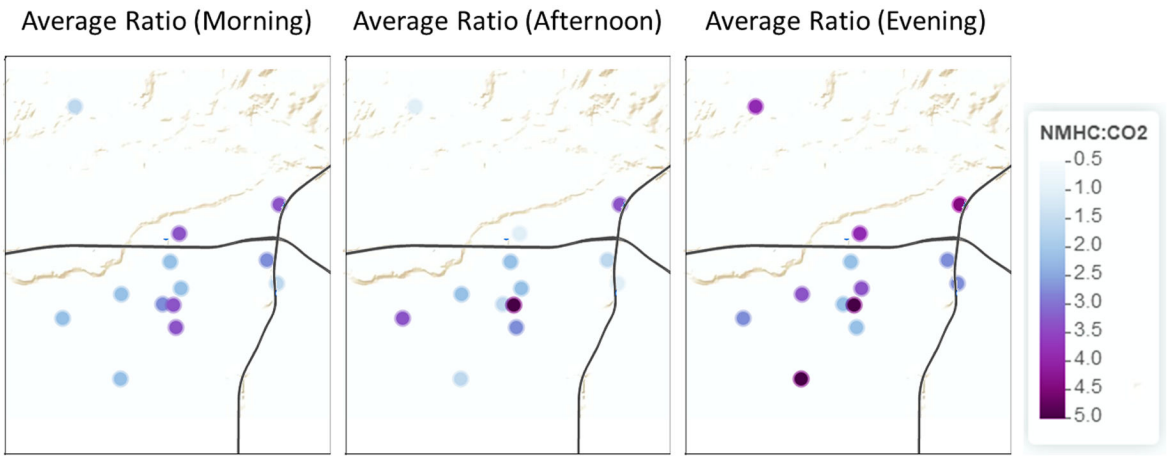

Figure 13:

Three maps of the average TNMHC:CO_2 ratio at each site in ppb/ppm for the baseline removed data sets. The map to the left, for the morning hours, depicts the average for 6 – 9am. The middle map depicts afternoon hours from 12 −3pm. The map to the right, depicts evening hours from 5 – 8pm.

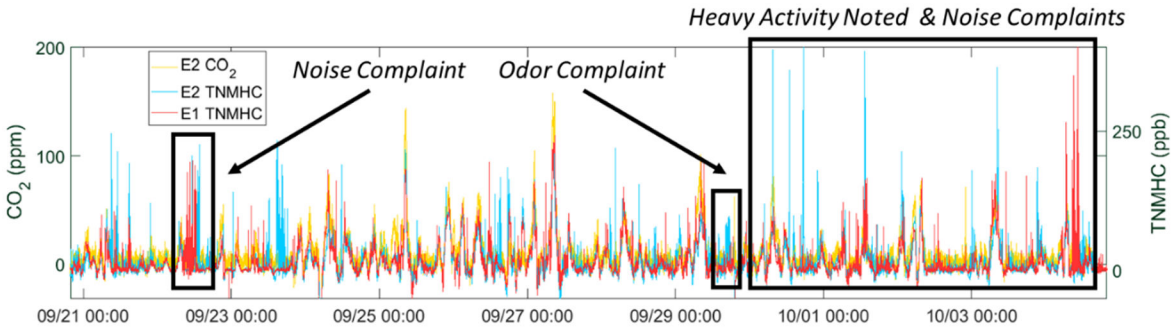

Figure 14:

Baseline-removed CO_2 and TNMHC data from Site E2 and Site E1, annotated with noise and odor complaints as well as observations by residents of heavy activity at the drill site.

Table 1:

Details of co-locations with reference instruments

Dates	Reference Instrument	Pollutants	Location
Pre: 8/1 – 8/4/2016	[a]Baseline Mocon Series 900 Analyzer	CH_4, TNMHCs	Los Angeles, CA
Post: 10/20 −10/28/2016	[a]Picarro cavity ring-down spectrometer	CH_4	Los Angeles, CA
Post: 11/2 – 11/12/2016	[b]Synspec Alpha 115	TNMHCs	Shafter, CA

[a]operated by SCAQMD,

[b]operated by SJVAD

Table 2:

Calibration Models

Model Name	Equation
Single Sensor CH_4	$C = \frac{Fig2600 - p_1 - p_3(T) - p_4\left(H^{-1}\right) - p_5(T_i) - p_7\left(T * H^{-1}\right) - p_8(T_d)}{p_2 + p_6(T)}$
Multi Sensor CH_4	$C = p_1 + p_2(Fig2600) + p_3(Fig2602) + p_4(Fig2600 * Fig2602) + p_5(T) + p_6(H) + p_7(T_i)$
Multi Sensor TNMHC	$C = p_1 + p_2(Fig2600) + p_3(Fig2602) + p_4(Fig2600 * Fig2602) + p_5(T) + p_6(H) + p_7(T_i)$

Predictors (lower case p with subscripts) are Fig2600 – R_S/R_0 for the Figaro 2600 sensor, Fig2602 – R_S/R_0 for the Figaro 2602 sensor, C – pollutant concentration, T – temperature, H – absolute humidity, T_i – continuous time, T_d – time of day. The predictor p_1 indicates an empirical constant.

Table 3:

Statistics from calibration model generation and validation, averages across 15 Y-Pods

	Training			*Testing*		
Target Pollutant	**R^2**	**RMSE**	**Mean Bias**	**R^2**	**RMSE**	**Mean Bias**
Single Sensor – CH_4 (ppm)	0.81	0.15	0.00	0.74	0.18	0.03
Multi-sensor – CH_4 (ppm)	0.88	0.11	0.00	0.80	0.16	0.07
Multi-sensor – TNMHC (ppb)	0.60	31.2	0.01	0.46	46.4	7.62

Table 4:

Statistics for the Co-location Period

Target Pollutant	Intra-Device R^2	Intra-Device RMSE	S/R (Y-Pod)	S/R (Ref)
Single Sensor – CH_4 (ppm)	0.98	0.04	6.0	5.9
Multi-sensor – CH_4 (ppm)	0.98	0.03	6.8	5.9
Multi-sensor – TNMHC (ppb)	0.97	6.48	1.2	0.7

Table 5:

Results of the generation and validation of models for the prediction of CO_2 and CO

	Training			*Testing*			Number of Y-Pods
Target Pollutant	**R^2**	**RMSE**	**Mean Bias**	**R^2**	**RMSE**	**Mean Bias**	
CO_2 (ppm)	0.91	7.72	−0.01	0.85	10.27	−4.43	15
CO (ppm)	0.88	0.088	0.0	0.87	0.05	0.01	5

Table 6:

Average correlation coefficient (R) between Y-Pod baseline data across varying spatial scales

	Single Sensor CH_4		*Multi-sensor CH_4*		*Multi-sensor TNMHC*	
Scale	**Complete**	**Baseline**	**Complete**	**Baseline**	**Complete**	**Baseline**
All deployed sensors	0.70	0.72	0.72	0.76	0.60	0.59
Sensors within neighborhood	0.81	0.85	0.74	0.81	0.72	0.75
Sensors (E1 and E2, ~140 m apart)	0.93	0.99	0.86	0.95	0.87	0.97

Table 7:

Coefficients of Determination between CO_2 and TNMHC signals for the events shown in Figure 8

	Panel A		***Panel B***		***Panel C***	
	R2 - CO_2	**R4 - CO_2**	**R2 - CO_2**	**R4 - CO_2**	**R2 - CO_2**	**R4 - CO_2**
R2 - NMHC	0.56	0.06	0.48	0.45	0.78	0.54
R4 - NMHC	0.01	0.68	0.39	0.81	0.71	0.89

Table 8:

Confusion matrix comparing resident observations and the occurrence of peaks at sites E1 and E2

	Observation Noted	No Observation Noted
Peaks Noted (>= 3 per day)	26	16
Few Peaks Noted (< 3 per day)	3	10
Rates	True Pos. = 0.90	False Pos. = 0.62
	False Neg. = 0.10	True Neg.= 0.38

Table 9:

Confusion matrix comparing resident observations and the occurrence of peaks at sites R3 and N4

	Observation Noted	No Observation Noted
Peaks Noted (>= 3 per day)	15	5
Few Peaks Noted (< 3 per day)	14	21
Rates	True Pos. = 0.52	False Pos.= 0.19
	False Neg. = 0.48	True Neg. = 0.81