Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors

Kira Sadighi[1], Evan Coffey[1], Andrea Polidori[2], Brandon Feenstra[2], Qin Lv[3], Daven K. Henze[1], Michael Hannigan[1]

[1]University of Colorado Boulder, Department of Mechanical Engineering, 427 UCB, 1111 Engineering Drive, Boulder, CO 80309, USA
[2]South Coast Air Quality Management District, Air Quality Sensor Performance Evaluation Center, 21865 Copley Drive, Diamond Bar, CA 91765-4178, USA
[3]University of Colorado Boulder Department of Computer Science, 430 UCB, 1111 Engineering Drive, Boulder, CO 80309, USA

Correspondence to: Kira Sadighi (Kira.Sadighi@colorado.edu)

Abstract. Sensor networks are being more widely used to characterize and understand compounds in the atmosphere like ozone (O_3). This study employs a measurement tool, called the U-Pod, constructed at the University of Colorado Boulder, to investigate spatial and temporal variability of O_3 in a 200 km^2 area of Riverside County near Los Angeles, California. This tool contains low-cost sensors to collect ambient data at non-permanent locations. The U-Pods were calibrated using a pre-deployment field calibration technique; all the U-Pods were collocated with regulatory monitors. After collocation, the U-Pods were deployed in the area mentioned. A subset of pods was deployed at two local regulatory air quality monitoring stations providing validation for the collocation calibration method. Field validation of sensor O_3 measurements to minute resolution reference observations resulted in R^2 and root mean squared errors (RMSE) of 0.95 – 0.97 and 4.4 – 5.9 ppbv, respectively. Using the deployment data, ozone concentrations were observed to vary on this small spatial scale. In the analysis based on hourly binned data, the median R^2 values between all possible U-Pod pairs varied from 0.52 to 0.86 for ozone during the deployment. The medians of absolute differences were calculated between all possible pod pairs, 21 pairs total. The median values of those median absolute differences for each hour of the day varied between 2.2 and 9.3 ppbv for the ozone deployment. Since median differences between U-Pod concentrations during deployment are larger than the respective root mean square error values, we can conclude that there is spatial variability in this criteria pollutant across the study area. This is important because it means that citizens may be exposed to more, or less, ozone than they would assume based on current regulatory monitoring.

1 Introduction

Tropospheric ozone formation and destruction is a complex chemical process involving a series of interdependent chemical reactions of volatile organic compounds (VOCs) and nitrogen oxides (NO_x) in the presence of ultraviolet (UV) radiation (Jacob, 2000). The reactants are produced and consumed both naturally and through anthropogenic activities, as well as through atmospheric chemical reactions. In urban areas, the sources of these emissions and their impact on ozone formation vary in time and space. For example, trucks and cars, acting as mobile sources of primarily NO_x and VOCs, respectively, contribute to the formation and/or destruction of ozone depending on mixing ratios of each and the presence of UV radiation. Due to the health implications of increased ozone exposures, local, regional and national regulatory bodies have the obligation to measure, report and mitigate ambient ozone levels according to the National Ambient Air Quality Standards (NAAQS) (EPA, 2013).

The equipment employed at air quality monitoring stations (AQMS) is relatively expensive (>$100k/station) and requires substantial resources to maintain (e.g., technical expertise, shelter, land and power). As such, increasing the spatial resolution of the AQMS network is not readily feasible. Thus, one benefit of low-cost, portable sensing technology is the ability to collect data at more locations, increasing spatial resolution of existing AQMS. These technologies typically range in cost of $1-5k yet often require significant data retrieval and processing resources in addition to extensive characterization of the sensor in a given application. These technologies, in virtually all applications, still depend on reference grade measurements or standards in order to fulfil most research objectives. As such, many view these tools not as replacements of regulatory measurements but rather a supplement to them (Clements et al., 2017). Detecting pollutant variability between the regulatory AQMS supports the idea that more detailed information can be obtained by increased monitoring between existing stations.

Regulatory monitoring for compliance with the ozone NAAQS is undertaken as dictated by the Code of Federal Regulations, which states, "The goal in locating monitors is to correctly match the spatial scale represented by the sample of monitored air with the spatial scale most appropriate for the monitoring site type, air pollutant to be measured, and the monitoring objective." (EPA, 2006). Ozone monitoring site types include: highest concentration, population oriented, source impact, general/background and regional transport, and welfare-related impacts. Siting involves choosing a monitoring objective, selecting a location that best achieves those goals, and determining a spatial scale that fits the monitoring objective.

The minimum number of ozone monitoring sites required by the US Environmental Protection Agency (EPA) via the Code of Federal Regulations (CFR) in the Riverside and San Bernardino counties is three, given the population is between four and ten million. As of 2013, there were 20 active regulatory sites measuring ozone in Riverside and San Bernardino counties (California Air Resources Board, 2013). While this monitor density is more than sufficient for regulatory requirements, recent studies suggest that the current spacing is not sufficient to capture high spatial resolution of concentration variations (Bart et al., 2014; Moltchanov et al., 2015). This variability could potentially be used to inform exposure assessment for health studies as well as improve our understanding of pollutant sources and fate (Simon et al., 2016; Lin et al., 2015; Blanchard et al., 2014).

Networks of air quality sensors have been deployed in various settings. Moltchanov et al. (2015) measured O_3, NO_2 and VOCs in Haifa, Israel in the summer of 2013 to test the viability of sensor networks measuring small scale (100s of meters) intra-urban pollution. Two of the sites used in that study, sites A and B, had correlations between 0.82 and 0.94 with each other, but correlations between A or B and a third site, C, were much lower, between 0.04 and 0.72. Their finding of spatiotemporal variability on a neighbourhood scale means that spatiotemporal variability on the scale of <10km can also be expected. This finding of spatial variability at that temporal and spatial scale was not linked with robust in-field sensor validation that would ensure the result was actual concentration differences instead of measurement artifacts. Sensor validation is an important component of using low-cost sensors because they are subject to drift and confounding species. Drift is the change in measured concentration with time because of factors inherent to the sensor, not necessarily the environment that is being measured. Many metal-oxide sensors have been found to be affected by high temperatures and humidity (Rai et al., 2017). In 2013, Williams et al. (2013) quantified a tungstic oxide ozone sensor in the lab while addressing some of the main drawbacks associated with metal oxide (MO_x) ozone sensors (i.e. drift/long term stability, material degradation and sensitivity fluctuations). The ozone sensors in that study were held in a temperature-controlled environment, as the tungsten oxide sensor's conductivity varies strongly with temperature and may affect the concentrations. In the work presented here, temperature was included as a term in the model in an effort to address this issue after, rather than before, data collection. Researchers also deployed these gas semiconductor sensors in British Columbia over

roughly 10,000 km^2 for three months finding low errors (3 ±2 ppbv) between hourly averaged sensor and reference instruments while documenting the challenges of using, in this instance, wireless sensor networks (Bart et al, 2014). Lin et al. (2015) demonstrated high correlations (0.91) between tungsten oxide semiconductor ozone sensors and hourly averaged Federal Reference Method (FRM) chemiluminescence gas analyzer measurements in Edinburgh, UK with similar magnitudes. While many of these studies show good agreement between metal oxide sensors and reference instruments; there is still a need for uncertainty estimation and framing of the deployment results in light of those uncertainties.

Here we specifically seek to answer the question, are these metal oxide sensors able to detect significant differences on scales that are smaller than current EPA reference stations, given their quantification uncertainty? This study is unique in that the Inland Empire region of the greater Los Angeles frequently experiences high levels of ozone resulting in nonattainment of the NAQQS ozone standard. The combination of abundant sunlight and high VOC concentrations in the presence of NOx is conducive for the formation of ozone. The Pacific inversion layer over southern California and mountains that form a natural basin act together to keep pollutants from dissipating (Littman and Magill, 1953). Moreover, the regional air quality regulatory body, South Coast Air Quality Management District (SCAQMD), has expressed increased interest in low-cost air quality sensor applications and recently installed the nation's first testing center for such technologies. As such, Riverside, CA is an ideal test bed to answer our research question.

2 Methods

This field study was conducted within a 200 km^2 area of northwestern Riverside county, California, a region frequently designated as nonattainment for failing to meet requirements for ozone and particulate matter designated by the EPA (EPA, 2016). Thirteen low-cost ozone monitors were deployed within an 8 km radius in Riverside in the summer of 2015 (Fig. 1). These monitors were sited in the cities of Riverside and Jurupa Valley with the aid of SCAQMD. Sites were chosen based on availability and power access. Ten locations were identified (Fig. 1) representing a variety of site conditions ranging from university campuses and residential neighbourhoods to commercial and industrial zones. Within this area, there are two regulatory AQMS that measure O_3: Rubidoux and Mira Loma. The transportation authority in California, Caltrans, records traffic volume information for many large highways. Annual average daily traffic (AADT) is recorded at many road intersections. On two major roads in the study area in this region, specifically Hwy 91 and Hwy 60, the averaging of all the milepost traffic count data between junctions shows AADTs of 180,500 and 220,500, respectively ("2015 Traffic Volumes", 2017). Van Buren Avenue does not have AADT data. However, it has two lanes each way, while the other highways have more than four. In general, there is a large number of vehicles traveling around and through this study area daily; these vehicles likely represent the dominant sources of NO_x, and VOCs, precursors to ozone formation.

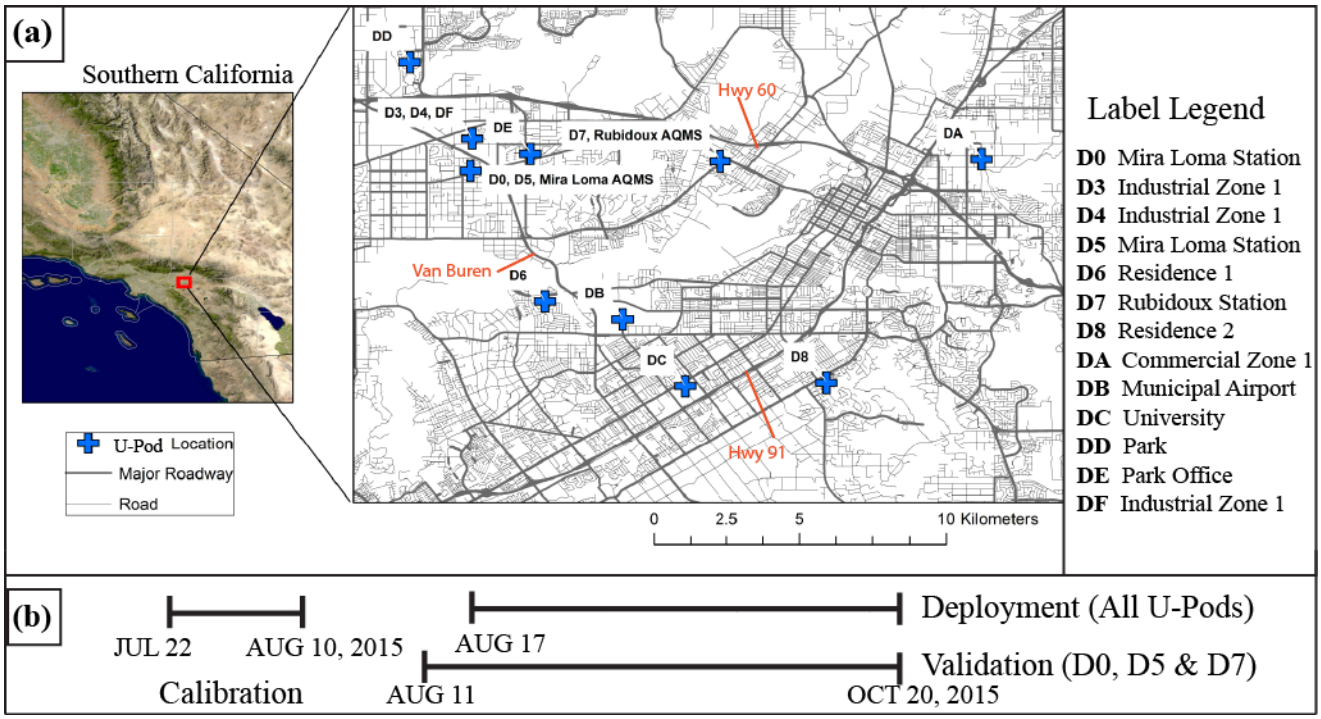

Figure 1. (a) A map of the deployment area. The crosses indicate U-Pod locations, with the AQMS labelled by name and (b) a timeline of project phases, from calibration to deployment. Validation overlapped with the deployment time period.

2.1 Low-Cost Monitor

Measurements were taken using the University of Colorado U-Pod air quality monitoring platform (mobilesensingtechnology.com) described in previous work (Piedrahita et al. 2011). Briefly, the U-Pod consists of an Arduino data acquisition system and a suite of environmental sensors enclosed in a small, ventilated, portable case (Fig. 2). Specifically, O_3 is measured using a metal oxide (MO_x) sensor, (MiCS 2611, SGX Tech. formerly e2v ~ $11). Enclosure air temperature and relative humidity were also measured. U-Pod locations were verified using an on-board GPS chip and all data were saved to a micro SD card. Logged data were collected into minute medians to match the highest temporal resolution of nearby regulatory air quality stations. Median values were used to reduce the influence of outliers within each minute. Duplicate O_3 sensors were included in most U-Pods to investigate sensor variability and model performance.

MO_x sensors operate through reduction/oxidation processes at the gas-semiconductor surface resulting in changes in electrical resistance (Barsan and Weimar, 2001; Korotcenkov et al. 2007). This change in resistance is in part a function of the concentration of the target gas (i.e., ozone) in the surrounding air, as well as temperature and humidity. Comprehensive reviews of MO_x gas sensors (Korotcenkov et al., 2007) and experimental tests (Masson et al., 2015; Rai et al., 2017) document potential concerns of using sensors in long term ambient monitoring campaigns and other sensing applications. A variety of environmental factors such as long-term exposure to water causing hydration of the oxide surface layer can lead to drift in the sensing chemistry, as well as cross sensitivity to other oxidizing species like NO_x. This poses special concern for conditions amenable to condensation. The MiCS 2611 datasheet warns specifically of overheating, a cause of sensor degradation or possibly permanent damage. Heating power supplied to the sensing resistor at 80mW is recommended to keep this element at 430°C (e2v MiCS-2611). Lower sensor resistor temperatures can result in decreased sensitivity and longer response times making measurements of heater element voltage and/or well-regulated circuits valuable in regards to long term sensor integrity (Masson et al., 2015). The magnitude and sources of sensor variability from this study are discussed further in Section 3.1.

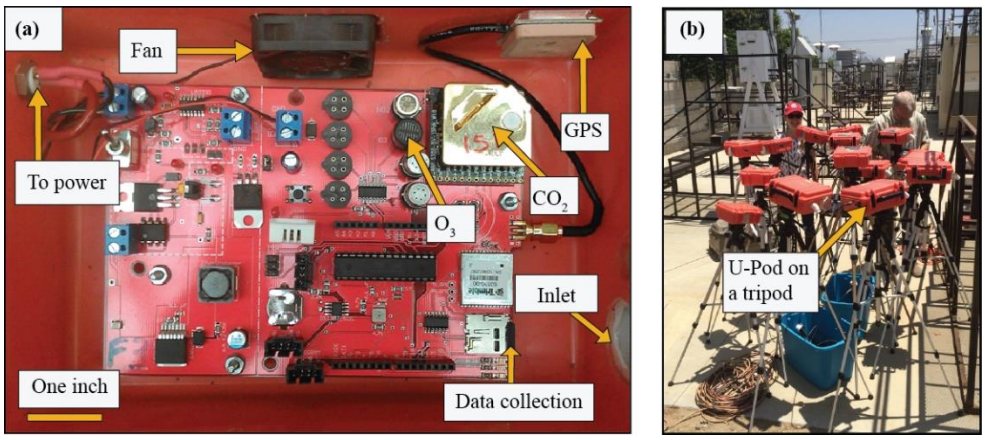

Figure 2. Demonstration of the U-Pod layout (a), including sensor locations and other features. (b) A photo of the field calibration collocation at Rubidoux AQMS.

2.2 Field Calibration

Sensors were calibrated using a field calibration technique commonly employed with low-cost sensor networks which involves collocating sensors with a reference grade monitor for an extended period of time prior to and/or directly following a field deployment (Piedrahita et al, 2011). The concept of field calibration is straightforward: develop regressions between the reference measurement and gas sensor signal using combinations of concurrently collected environmental data. All U-Pods were calibrated at the SCAQMD Rubidoux AQMS (elev. 248m above sea level) for three weeks, July 22 – Aug 10, prior to the field deployment. The Rubidoux station sampling scale is classified as "urban" for ozone and is located 119 m from Hwy. 60 (SCAQMD, 2015). Reference ozone is measured using a designated Federal Equivalent Method (FEM) Thermo 49i dual cell UV photometric monitor. This monitor is equipped with temperature and pressure compensation, which adjusts for changes in sensor signal due to changes in the sample gas. Numerous field calibration relationships were developed using a suite of custom MATLAB codes. This process involves performing linear and nonlinear regressions using sensor signal, measured U-Pod enclosure temperature, absolute humidity and time (to account for sensor drift) against the reference gas concentrations. MO_x sensor signals are the ratio of instantaneous resistance to a reference resistance defined during the field calibration. To evaluate the resulting regression fit, we used coefficient of determination (R^2), root mean square error (RMSE) and explored residuals with relation to each input variable, specifically looking for normal distributions. An interaction term between temperature and ozone concentration improved the model fit at higher mixing ratios leading to overall higher correlations, lower error, and improved residual distributions (see Table 1 in Section 3). The best performing model for ozone during calibration incorporates temperature, absolute humidity, and time, and is also referred to as the linear 4T model (Eq. 1).

$$S = p_1 + Cp_6(T + p_2) + Tp_3 + Ap_4 + (t - t_o)p_5 \quad (1)$$

In Equation 1, S is the sensor signal in R/R_o, where R is the sensor resistance and R_o is a specific normalizing resistance value. C is the pollutant concentration in ppbv, T is the temperature in Kelvin, A is absolute humidity in mole fraction, $t\text{-}t_0$ is the duration since the start of the calibration and the p variables are coefficients determined by the regression minimising least squares. Throughout this paper, concentration refers to the ozone mixing ratio. In this model, a global absolute humidity term was employed;

this absolute humidity was calculated using Rubidoux reference station temperature and relative humidity, and a constant pressure, and was used in all U-Pods throughout the measurement campaign. The values of these coefficients are described in section 3.1.

2.3 Field Deployment

Following the field calibration, the U-Pods were relocated throughout the study area to the sites shown in Fig. 1. Sites were chosen based on availability and zoning. A mix of industrial, residential and commercial areas were selected including a university campus and public parks. U-Pod D7 remained at Rubidoux station while D0 and D5 were relocated to Mira Loma Reference station for the purpose of validation.

2.4 Field Validation of Model Performance

To quantify the performance of the calibration model coefficients, a nearly three month long validation dataset was collected comparing reference grade gas concentration measurements to sensor data after applying the model coefficients to the raw sensor data. Previous air quality sensor campaigns have either had mixed results when performing validation in the field or no validation was included. Moreover, no study, to our knowledge, has validated ozone sensor measurements to reference grade monitors at one-minute resolution. Two validation approaches were investigated. First, we compared sensor measurements to reference grade observations in the *same* location as was used for the field calibration. Second, we compared sensor measurements to reference grade observations in a *different* location from the field calibration site. The second approach can be used to address error associated with site specific confounders, such as NO_x or transient temperature effects present away from the initial collocation site. U-Pod D7 was validated using the first approach, as it remained at Rubidoux AQMS for the duration of the deployment. U-Pods D0 and D5 were moved from Rubidoux AQMS, after the calibration, to Mira Loma AQMS and validated using the second approach. The outcome of the field validation is presented in the results.

3 Results

3.1 Field Calibration Results

Calibration results for various models showing correlation and RMSE of the calibrated ozone data against the reference monitor data are provided in Table S1. For the sake of simplicity, results from the overall best performing model, see Eq. 1, are shown in Table 1. R^2 values and errors (RMSE) range from 0.97 – 0.99 and 1.8 – 3.9 ppbv, respectively.

Table 1: Field calibration results of the model, see Eq. 1, for ozone sensors showing R^2 and RMSE with the reference monitor data. Two O_3 entries means there are two different sensors in the same U-Pod.

U-Pod ID	D0	D3	D4	D5	D6	D7	D8	DA	DB	DC	DD	DE	DF
Sensor 1 R2, RMSE	0.98, 3.1	0.98, 3.0	0.98, 2.6	0.99, 2.7	0.98, 3.5	0.98, 2.8	0.98, 3.0	0.97, 3.9	0.98, 2.8	0.99, 2.6	0.99, 1.8	0.97, 3.4	0.98, 3.1
Sensor 2 R2, RMSE	0.98, 3.2	0.98, 3.0	0.98, 2.7	0.98, 3.0	0.99, 2.4	0.98, 3.0		0.97, 3.9		0.98, 2.7	0.99, 1.8	0.98, 2.9	0.98, 3.0

Figure 3 illustrates the calibration results for U-Pod D0. Residuals were calculated as modeled minus reference instrument concentrations. The normally distributed residuals shown in panel c were indicative of an unbiased model. Residuals were plotted versus various model parameters to assess bias in the model performance as a function of the predictors. The slightly negative slope of the trend line in panel e indicated under predicting at increasing absolute humidity whereas positive slopes in panels d and f shows the opposite trend, slight over-prediction at higher values of concentration and temperature. The R^2 and RMSE values for the calibration of this sample U-Pod were 0.97 and 2.9 ppbv respectively.

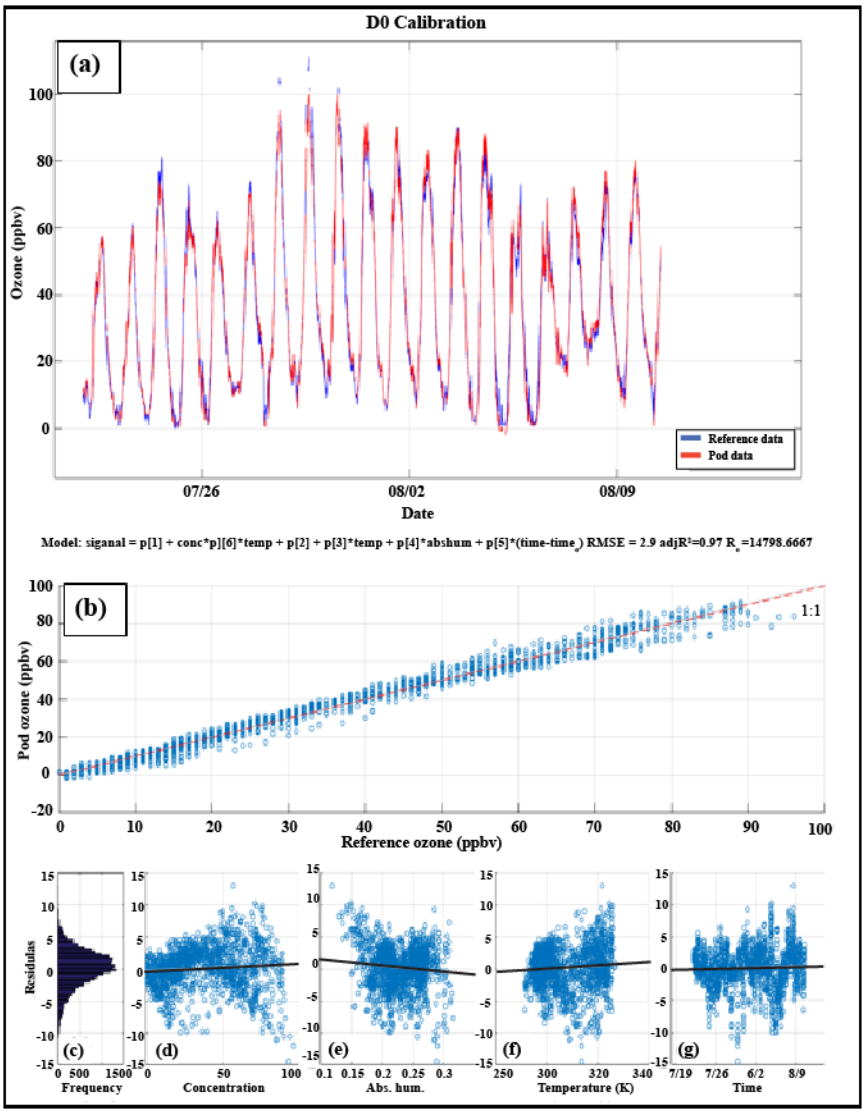

Figure 3. Example calibration results for one ozone sensor in U-Pod D0. Panel (a) shows the modeled ozone sensor time series (red) with the reference measurements (blue) along with the model expression below and (b) shows a scatterplot of the minute measurements, (c) the distribution of residuals and the relationship between residuals and model variables: (d) concentration, (e) absolute humidity, (f) temperature, and (g) time.

The quickly expanding sensor community has been convening to discuss practical and theoretical considerations of low-cost sensor applications in the modern landscape identifying a need for increased understanding of inter-sensor variability (Clements el al., 2017). Few groups have thoroughly investigated the physiochemical relationships governing MOx (and more specifically tin oxide) sensor operating principles. Yet, Barsan and Weimar (2001) and subsequently Masson et al. (2015) lay forward an in-depth discussion on MOx conduction models and how those models incorporate chemical kinetics and semiconductor electrical

properties in explaining sensor signals. Masson et al. focused particular attention to temperature effects finding ambient temperature to be one of the most significant confounders in ambient air monitoring using CO sensors (MiCS-5525). Petersen et al. explored the experimental effects of power supply fluctuations on O_3 (MiCS-2614) and NO_2 (MiCS-5914) sensors as it relates to acute sensor response and long term sensor stability finding different responses from sensors exposed to the same environment – crediting these differences to mainly manufactural discrepancies (Petersen et al., 2017).

Additional insight into this effort can be gleamed by exploring the results of sensor-specific model parameters from the nearly three-week calibration period of this study. To directly compare model parameters (i.e., coefficients), standardized regression coefficients were generated by rescaling model input variables from 0 to 1. Rescaling was achieved by dividing the difference between each variable data point from its respective distribution minimum by the maximum difference measured (e.g., $[v_i - v_{min}]/[v_{max} - v_{min}]$). This process allows one to directly compare the magnitude of one predictor variable to any other; an advantage of dimensionless analysis. Figure 4 shows the fractional contribution of each model parameter during the calibration period towards estimating the sensor signal (R/R_o). Concentration (reference, ppbv) and the concentration temperature interaction term combined explain 86% of the predictive capability of Eq.1 for the average sensor used in this campaign. The temporal drift coefficient (p5) contributes less than one percent to the overall regression indicating minimal signal drift during the 19 days of calibration and also explaining the minimal improvements in the descriptive statistics from the "Linear 3" and "3T" models to the calibration models including a temporal drift term (e.g., "Linear 4" and "4T", see Table S1). Absolute humidity, temperature and the intercept, combined, are less than 15% of the total predictive contribution. Figure 4 acts as visual evidence as to the significance of the concentration-temperature interaction feature in this sensor model and perhaps other gas-specific MOx sensor models. This interaction term could be capturing what Masson et al. discovered when performing MOx sensor signal regressions with temperature and CO reference gases; namely, "this improvement of fit with concentration coincides with the observation that the response data [R/R_o] becomes more linear with temperature as concentration is increased" (Masson et al., 2015). Figure S1 illustrates the inter-sensor standardized regression coefficient variability.

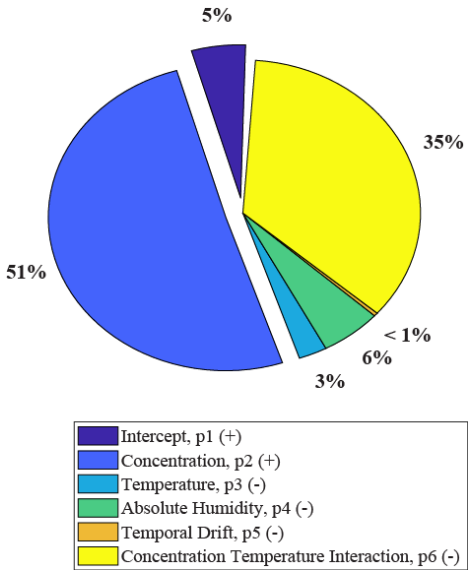

Figure 4: Average relative effect size of model parameters predicting sensor signal (R/R_o) from standardized regression coefficients. The direction of the parameter effect is shown in the legend (+ or -).

It is important to note that the reference resistance, R_o, which is the resistance in clean air, had moderately high inter sensor variability; a coefficient of variance (standard deviation divided by the mean) of 0.92. This reference resistance corresponds to the

minimum resistance at 25 °C, and each sensor has a different R_o. Differences in R_o could possibly be explained by sensor age or even MO_x nanostructure as posed by some research (Sun et al., 2012). Manufacturer heterogeneity, sensor age and lifetime exposure to oxidants are posed as potentially contributing to this variation but more investigation is recommended in future sampling (Rai et al., 2017).

3.2 Deployment Data Filtering and Processing

Some temperature and humidity values were experienced by the U-Pods during the deployment that were not experienced during the calibration time period. This means that the environmental parameter space sampled during the calibration time did not cover the parameter space experienced during the deployment. Deployment data were filtered for conditions that would require extrapolation, an example of which is shown in Fig. 5. Because ozone measurements are dependent on temperature and humidity, one way to reduce error in the deployment data is to only use ozone data points whose temperature and humidity were in range of that of the calibration data. All U-pod data from the deployment period were filtered to eliminate points that had temperature and relative humidity values out of the ranges recorded during calibration. The global absolute humidity in Fig. 5a is the same for all U-Pods. Normally, the absolute humidity would be calculated for each U-Pod using its individual recorded temperature, relative humidity, and pressure. However, during the deployment, the relative humidity sensors failed in several U-Pods. The relatively high chance of sensor failure in the field is one of the limitations of low cost sensor networks. Four of the U-Pods experienced RH values below zero. However, the RH sensor sets these values to zero. Therefore, there was no way to recover any data below zero. All of the U-Pods experienced, at some point, at least one week of missing data. Because of this, temperature and relative humidity data from Rubidoux AQMS, along with a constant pressure value, were used to calculate the global absolute humidity for the Riverside area for each minute. During calibration, the same values of absolute humidity were used for each U-Pod, but temperatures were U-Pod specific.

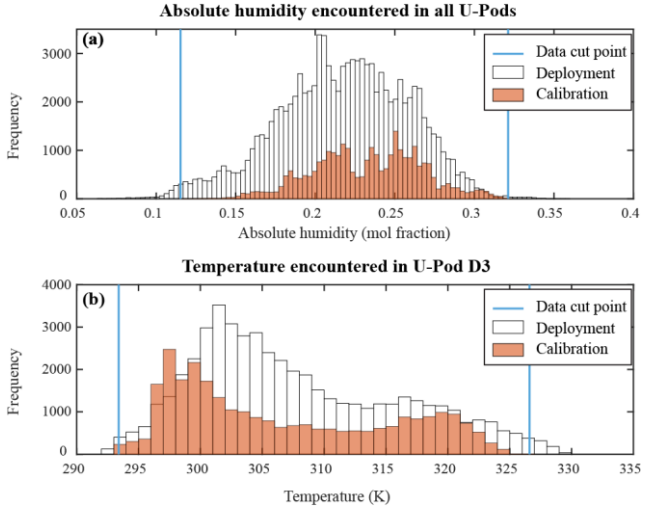

Figure 5. Example filtering for a U-Pod (D3) showing lower absolute humidity (a) and higher temperatures (b) occurred during the deployment than during the calibration. The data cut point shows where minimum and maximum values of the variables included in the data were excluded.

In addition, deployment data were filtered for maximum values of O_3. In some instances, the ozone data spikes to unrealistically high levels. The 95th percentile of the absolute differences between the two reference stations during the calibration period was 11

ppbv. The maximum one-minute value recorded by either station during this time was 160 ppbv. As such, we employed 171 ppbv as a realistic maximum level of ozone to expect across the study area. Concentrations that were over this threshold were removed. No minimum filtering was needed for O_3.

Lastly, data were filtered using consecutive differences. Data were omitted when they fell more than eight standard deviations away from the mean consecutive difference in values. This is a standardized way to cut out spikes in data caused by power control issues. The results of the deployment data filtering, including percent of data lost, are shown in Table S2. Most U-Pods (except D8 and DB) have two ozone sensors. For U-Pods with two ozone sensors, only one was used for the analysis. The data from the calibration time period for each sensor was compared to the reference data at Rubidoux. Whichever sensor had the highest correlation and lowest RMSE with the reference was chosen for subsequent analysis.

U-Pod DD was omitted from this analysis due to a lack of data. This pod lost almost 46% of its data after the filtering process and collected significantly less data than the others due to site security issues. U-Pods D4, D5, D6, D8 and DF required a modification be made to their electronics boards. This modification to the U-Pod system appeared to have shifted ozone baseline signal values resulting in biased values for D5 (see section 3.3 below). In a conservative effort, all U-Pods that were modified as described above were removed from the subsequent ozone analysis. Since some U-Pods were at the same location, the removal of these U-Pods resulted in the loss of three sites from the study. All the remaining sites were left with one U-Pod each.

3.3 Validation of Field Calibration

Validation of the field calibration models was achieved by deploying U-Pods next to reference instruments during times when the others were spread out over the study area. The validation time period (Aug 11 – Oct 25) overlapped with the deployment time period (Aug 17 – Oct 20). Coefficients generated from the regression models (Table S1) were applied to the filtered data from D7, D0 and D5. The best performing model was selected based on R^2, RMSE and residual distributions. Ozone concentrations were best modeled over the entire validation time period using the model shown in Eq. 1, similar to what was observed for the calibration. The purpose of this comparison was to verify that the model that resulted in the best statistics for the calibration, also did so for the deployment time period. In order to gain a better understanding of the dependency of model performance on the selection of the validation data, we randomly selected 10% of the validation data and calculated validation statistics for this subset of the validation period and repeated this process 200 times. This iterative method allows us to assess the sensitivity of the validation statistics to the data randomly selected. The resulting distributions for the performance metrics are shown in Table 2. Tight distributions show little dependence on the data selected. Detailed results from the entire validation period are presented in Figs. S2, S3 and S4 for pods D0, D5 and D7, respectively.

Table 2. Overall validation sensitivity results showing mean residuals, median residuals, R^2 and RMSE of sensor measurements against Rubidoux or Mira Loma AQMS O_3 (ppbv) observations. Two-hundred iterations of 10% randomly selected minute-data were used for validation statistics (± 1 SD).

U-Pod ID	mean residual	median residual	mean R^2	mean RMSE	validation method
D7 O_3 Sensor 1	2.4 ±0.1	1.2 ±0.1	0.965 ±0.001	5.6 ±0.1	Same location
D7 O_3 Sensor 2	2.8 ±0.1	1.5 ±0.1	0.963 ±0.001	5.9 ±0.1	Same location
D0 O_3 Sensor 1	0.7 ±0.1	0.8 ±0.1	0.974 ±0.001	4.4 ±0.1	Different location

D0 O_3 Sensor 2	1.1 ±0.1	1.0 ±0.1	0.971 ±0.001	4.9 ±0.1	Different location
*D5 O_3 Sensor 1	5.5 ±0.1	5.1 ±0.1	0.971 ±0.001	5.0 ±0.1	Different location
*D5 O_3 Sensor 2	6.4 ±0.1	3.9 ±0.1	0.953 ±0.001	7.2 ±0.1	Different location

*D5 experienced an electrical issue resulting in data omission from analysis

The first validation method (U-Pod in the same location as the reference station, D7) would be expected to have better validation statistics than U-Pods validated using the second method (U-Pod relocated to a different location, D0 and D5) because the environmental conditions (e.g., temp, humidity, distance to roadway and other site-specific conditions) encountered by the pods were the same as the reference for the first validation method. However, in viewing the statistics, this is not the case as both O_3 sensors in D0 show better or similar performance to the Mira Loma station reference data than the two sensors in D7 compared to the Rubidoux station reference concentrations. For transparency, validation results from D5 were presented in Table 2 to show the effect of the electrical modification; the mean residuals for D5 are biased at 5.5 and 6.4 ppbv and much higher than those from D7 and D0. The mean RMSE from D0 and D7 sensors in Table 2 can be equated to the overall U-Pod uncertainty for the deployment.

Organizations using or planning to use sensors to monitor ambient air quality are interested in how frequently sensors require calibration as to keep them within a specified "tolerance" of reference-grade measurements. As a precautionary note, durations between suggested calibrations are highly dependent on the environment, quality and robustness of the calibration, and gas species of interest. The validation statistics presented so far have been aggregated over the entire deployment period (or have been selected at random) in the case of the iterative validation described above. However, to further inform the sensor community on how robust calibration models can be through time and environmental space (e.g., humidity and temperature), validation was performed independently for the first week and last full week of the deployment and the results for each week are shown below in Figure 6.

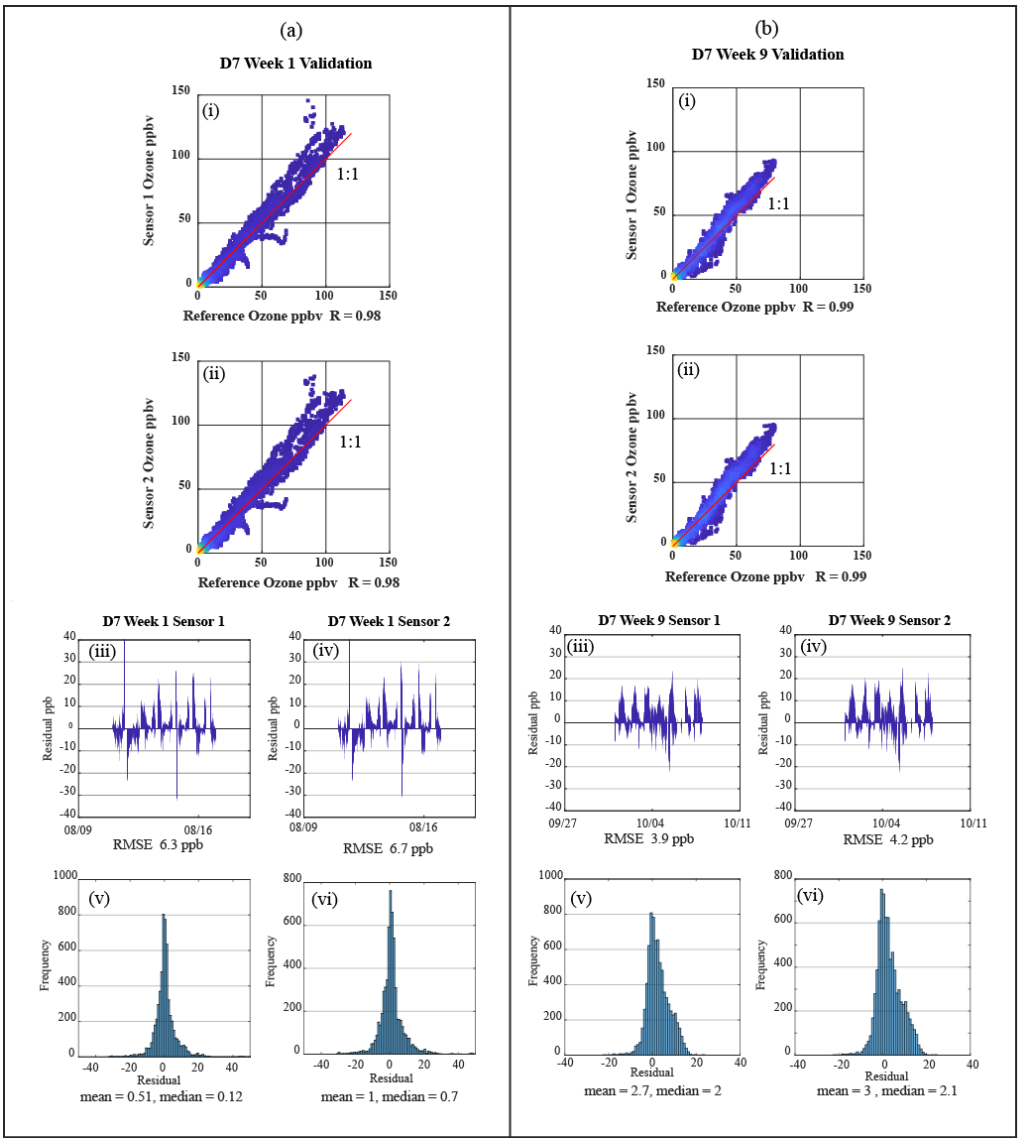

Figure 6: Validation results from the (a) first week and (b) ninth week of the deployment period for D7 ozone sensors separated by the red line. Subpanels (i) show a scatterplot of sensor 1 and reference measurements with warmer shading showing a higher density of points, (ii) show a scatterplot of sensor 2 and reference measurements with warmer shading showing a higher density of points, (iii) Depict residuals over time for sensor 1 with RMSE, (iv) depict residuals over time for sensor 2 with RMSE, (v) is a histogram of residuals with mean and median residual for sensor 1, (vi) is a histogram of residuals with mean and median residual for sensor 2.

Within the first week of the validation (panel a), the range of reference ozone concentrations (~0 to 115 ppbv) is much larger than those found in week nine (panel b, ~0 to 80 ppbv) although the Pearson's correlation coefficients (R) are remarkably high (≥0.98) for both sensors in both weeks (i, ii). The red lines are 1:1 lines, not lines of best fit. The residuals plotted as a function of time over each week (iii, iv) are similar in magnitude but by week nine (b; v-vi) there is a slight bias (mean = 2.7-3.0 ppbv) towards higher sensor measurements even though the RMSEs are lower in week nine (3.9 and 4.2 ppbv) than in week one (6.3 and 6.7 ppbv). Calibrations performed more frequently than every 9 weeks may reduce slight shifts in mean residuals. Monthly calibrations could balance monitoring resources and quality of ozone sensor data for a region like Riverside, but should be done on a case-by-case basis.

Figure 6 has two identifiable deviations from the 1:1 line. These two events, identifiable as the "claws" in week one (shown in panel (a (i-ii)), demonstrate higher reference measurements than both D7 sensors leading to large residuals. These claws are separated in time but each claw is a single event (consecutive measurements) lasting one and eight hours in duration. To explore these claws further, a scatterplot for each sensor colored by temperature and humidity at each time point were created (Fig. S5). They show that the two events visible for D7 occur at drastically different temperatures and humidity. The first (lower) claw has low temperature and high humidity, and the second has the reverse conditions. This finding provides evidence for a separate confounding variable, as it is not the same condition in temperature or humidity that causes these under predictions in ozone measurements. In future studies, the U-Pod could be outfitted with sensors to detect other possibly confounding gasses, such as NO_x or VOCs.

SCAQMD performed nightly precision checks (PC) consisting of measuring the ozone concentration of a known gas standard that typically ranges between 90-100 ppbv for one hour. When PC measurements deviated more than 5% from expected values (corresponding to approximately 5 ppbv), subsequent data would be flagged and a work order would be generated for service or calibration. Values that are within 5% of the standard would not be flagged. This serves as a reference point for the quality of the reference ozone measurements. During validation, O_3 sensors had measurement error (RMSE), median residual and mean residual ranges of 4.3 – 7.3, 1.7 – 5.2, and 0.6 – 6.5 respectively. Both median and mean of the residuals were calculated to assess bias. As discussed earlier, D5 experienced an electrical issue during the calibration period which resulted in a clear bias throughout the validation dataset. This particular electrical issue points to the challenges of using such sensor platforms in an ambient monitoring context, a topic widely discussed in the air sensor community (Kumar et al. 2015). Median bias for the other U-Pods was relatively small and on the order of 1-2 ppbv.

3.4 Deployment Data

As mentioned above, U-Pods were deployed, spread out across 200 km^2 area in Riverside, CA; as such, the aim of our data analysis is to present spatial differences of U-Pod measurements that include measurement uncertainty, and thus allow us to understand the ability of the sensors to detect variability. To examine this spatial variability, we computed the R^2 values and median absolute differences for all possible U-Pod pairs. Unless otherwise stated, median minute time resolution data recorded during the approximately 10 week deployment were used in the following analysis. The model coefficients obtained during the calibration time period (collocation with the reference monitor) were applied to all data during both the calibration and deployment time periods. Applying the model to the data collected during the collocation yields the best possible accuracy of the U-Pod sensors, as the model is being applied to the data from which it was derived. As such, comparisons of deployment data to collocation data are useful to assess the variability observed when the U-Pods are deployed vs. when they are collocated. This allows us to observe actual spatial and temporal differences. In all following figures, hours of the day are given in local time.

The U-Pods sampled for approximately 2900 hours total, 58% of which consisted of the deployment period data. The medians of ozone value distributions during the calibration range from 29-30 ppbv. During calibration, the 5^{th} and 95^{th} percentiles ranged from 2-5 ppbv and 70-83 ppbv, respectively. During deployment, the median ozone values were between 14 and 31 ppbv while the 5^{th} and 95^{th} percentile ranges were 0-6 ppbv and 67-99 ppbv, respectively.

Ozone concentrations experience a diurnal cycle. This cycle usually incorporates low ozone at night and during the early morning, and a peak in concentration sometime during the day. Gao (2007) used hourly ozone measurements recorded over southern

California from June 16th to October 15th, 1997 and found that ozone began to increase in the region around 8:00, peak between noon and 15:00, and then undergo reduction until about 21:00. The precursors to forming ozone: sunlight, VOCs and NO_x also have daily cycles, that in turn affect the ozone cycle profile (Gao, 2007). Figure 7 shows the diurnal cycle for ozone based on concentrations collected during this study.

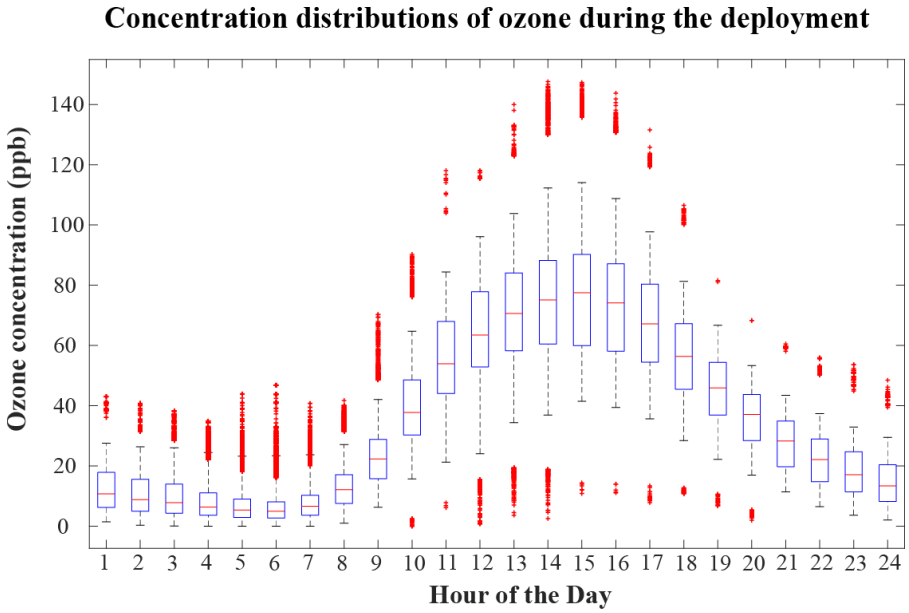

Figure 7. The diurnal cycle of ozone during the deployment. Distributions are concentrations from all U-Pods during each hour. Whiskers indicate the 5th and 95th percentile, with + marks falling outside of this range. The box boundaries span the 25th to 75th percentiles.

Figure 7 offers context of what the temporal variability in ozone concentrations in this study looks like. There are trends in ozone concentrations across Southern California that would be expected. Ozone is lowest from midnight to 6:00. Then the accumulation period takes place between 6:00 and 14:00. Peak concentrations occur between 14:00 and 16:00, and for the remaining hours, concentrations decrease again.

In order to assess spatial variability, we examined the R^2 values for all possible U-Pod pairs for each hour of the day. The larger the spread and smaller the magnitude of the R^2 values, the more spatial variability was likely present in that hour across the study region. Figure 8 shows correlation information between U-Pods for each hour of the day for ozone. For this plot, all data were binned by hour. Then within those bins, correlations were performed for every possible U-Pod pair. As such, each boxplot consists of 21 points.

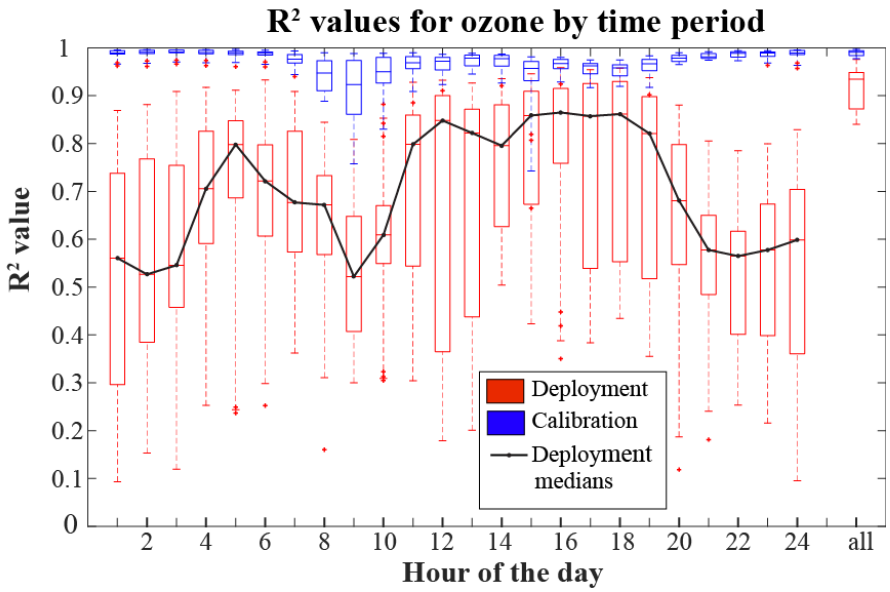

Figure 8. Each boxplot is a collection of the R^2 values between every pair of U-Pods for each hour of the day. There are 21 points in each boxplot. Medians of distributions are marked by horizontal lines. Whiskers indicate the 5th and 95th percentile, with + marks falling outside of this range. The box boundaries span the 25th to 75th percentiles. The "all" category includes all hours of the day.

U-Pod ozone measurements are more correlated to each other during calibration than deployment. The R^2 values between collocated pods are very high, with their medians varying from 0.92-0.99 ppbv. Conversely, spatially distributed pods were less correlated with each other, leading to R^2 distribution medians between 0.52 and 0.86. The "all" category in Fig. 8 represents the R^2 values between U-Pods, without binning by hour. The medians for the calibration and deployment in this column, respectively, are 0.99 and 0.93 ppbv, with slightly more skewness towards lower R^2 in the deployment distribution. It is only when binning by hour that greater differences are seen. U-Pods are most different from each other during the hours from 21:00 to 3:00, and at 9:00. U-Pods are most similar around 5:00 and between 11:00 and 19:00. Relationships in R^2 values between pods are changing most quickly through time between 3:00 and 11:00, and again between 19:00 and 21:00.

Absolute O_3 concentration differences between pairs of U-Pods were also examined to understand temporal and spatial variability. Figure 9 shows distributions of median absolute differences. All the minute median data were time-matched and binned by hour. Hourly datasets were paired to include every possible U-Pod pair. Within the time matched pairs, the median absolute difference between the two U-Pods was calculated. The distributions in Fig. 9 consist of those 21 points for each hour. The median values of these boxplots increase during the middle of the day, with two major increases observed at hours 10:00 and 15:00, and were lower during the night and early morning.

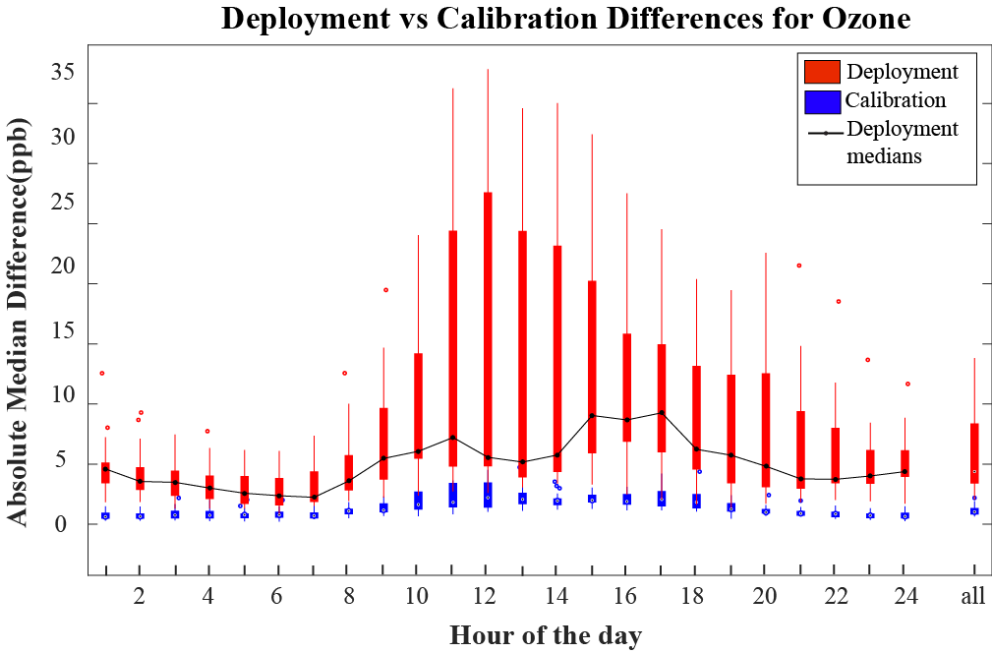

Figure 9. Distributions of medians of absolute differences between all pairs of pods for each hour of the day. Whiskers show 95% intervals. The black line connects the medians of the deployment. The "all" category includes all hours of the day.

We expected that times of day where the spatial variability was the lowest (R^2 highest) the smallest values of absolute differences would be observed. In other words, the deployment medians in Figs. 8 and 9 were expected to have an inverse relationship. There is an increase in R^2 while there is a decrease in absolute median differences around 4:00 to 5:00. There is also an increase in the differences that correspond to increasing R^2 with a peak around 9:00. The absolute median differences reach their minimums and maximums later than the R^2 values reach theirs by a few hours. Sometimes however, this inverse relationship between large R^2 and smaller differences does not appear. The second jump in median absolute differences between 15:00 and 17:00 was not reflected in reduced R^2 values during those same hours. From 6:00 to 10:00, the slope for the deployment medians in Fig. 9 is steep, indicating that pod differences were increasing quickly across the region, and over that same time period the spatial correlation was lower. The slope between 13:00 and 15:00 looks similar, but the R^2 values were roughly stable and relatively high. In other words, we observed spatial concentration differences and low correlation during the morning commute times, but in the afternoon when we observed the maximum concentration differences, we also observed relatively high spatial correlation. Absolute differences are growing during the morning period and into the afternoon, but since the whole area is experiencing accumulation, there is an increase in correlation as well. Furthermore, although Fig. 7 shows high concentrations during the day, Fig. S6 demonstrates that percent differences at these times are lower.

Towards the end of daylight hours, between 16:00 and 20:00, the medians of absolute concentration differences have a decreasing trend in time of day, which should be indicating that the U-Pods are becoming more similar because their differences are smaller. However, in the same hours and later, the R^2 values between all U-Pods decrease over time and remain low during the night, indicating that U-Pods are more different from each other than during the afternoon. Some studies have assumed negligible ozone precursor spatial differences in the first hours of the day and therefore spatial ozone homogeneity during the early morning hours (Moltchanov et al., 2005; Jiao et al., 2016). Figure 9 shows that the range of spatial absolute differences in O_3 is smallest at night. However, Fig. 8 suggests that spatial correlation at night is relatively low, causing concern for assumptions about the homogeneity of ozone concentrations at night for this location, although this assumption could be valid for other areas (Moltchanov et al., 2015).

Furthermore, the discrepancy between low absolute differences, but also low R^2 values may show that correlations alone are not enough to determine how similar two sites are. The actual differences in concentrations can reveal elements of spatial variability not captured by correlations, especially since correlations can be influenced by leveraging fewer high data points.

To further understand the factors impacting the observed spatial variability, we examined U-Pods individually in more detail. We undertook this investigation by comparing each U-Pod to a common reference U-Pod, to illuminate differences between locations in a normalized way. If no spatial variability was observed, then comparing two U-Pods' ozone measurements would show a 1:1 relationship with spread near the RMSE values determined in the validation (4.4-5.9 ppbv). To explore this analysis, D7 was used for normalization. U-Pod D7 was never moved from Rubidoux station throughout the project and as such was employed in the validation effort mentioned previously. This U-Pod was used as the normalization instead of an AQMS reference monitor in order to compare two similar types of measurement. The U-Pod to U-Pod comparisons are shown with the differences between calibration period trends and deployment trends in Figure 10 as well as hourly patterns in Figure 11.

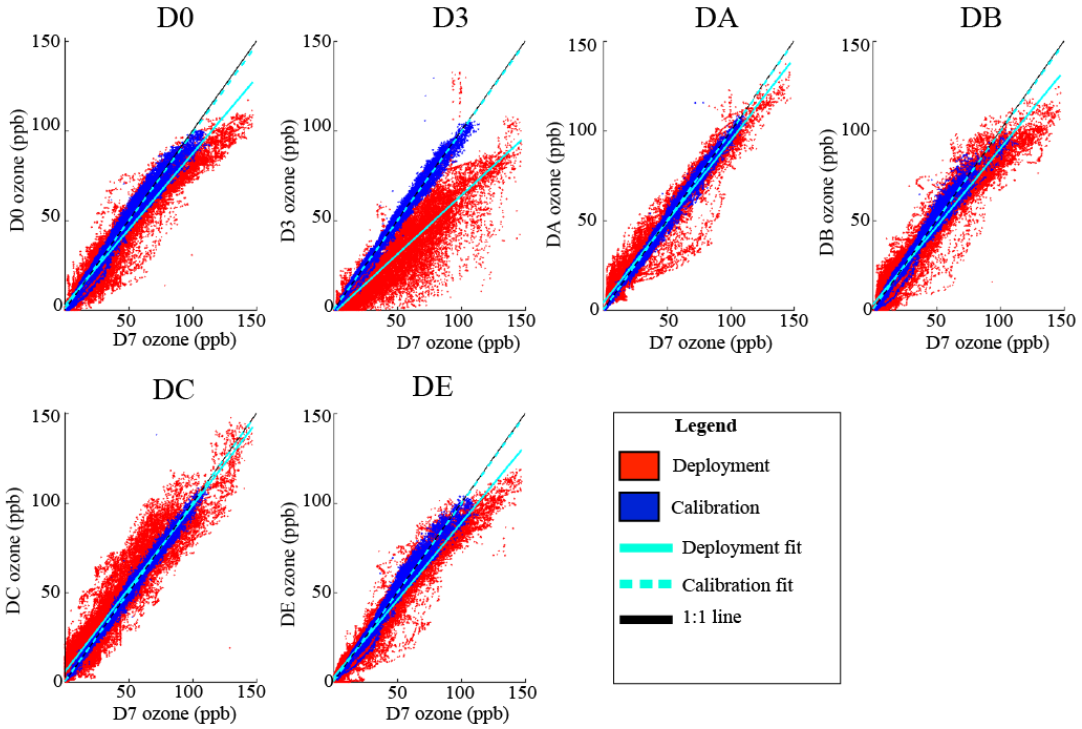

Figure 10. U-Pod D7 ozone concentrations are plotted on the x-axis and other U-Pod ozone concentrations recorded at the same times are on the y-axis. The sets are color coded according to time period their data were taken, and each color is fit with a linear line.

In Fig. 10, the calibration data points, representing collocated O_3 measurements, are consistently more densely grouped than the red data points which show the spatial deployment data. This further demonstrates that individual U-Pods were observing spatial differences in O_3. Also, D0, DA, DB, and DE have interesting deviations of O_3 concentrations away from the central cloud of deployment points, in the form of curved areas away from the center line. The deployment trend line slopes (solid line) are lower than the calibration slopes (dotted line). As such, D7 at the Rubidoux site typically measured higher O_3 than the other U-Pods that were spatially deployed (excluding DC and DA).

Examining the data in this way allows for detailed comparison of U-Pods at different sites. For example, sites D0, D3 and DE were not more than 1.8 km away from each other, near Van Buren Blvd. in the north west of the project area, and all were less than 1.2 km from the road. Therefore, one might expect data from these U-Pods to be very similar. Indeed, D0 and DE have similar data cloud shapes in Fig. 10. However, data from D3 looks to be rather different. This could indicate that a localized source is affecting the ozone concentrations at that site. Perhaps a local emission of NO was scavenging ozone at Industrial Zone 1 as a result of industrial operations. Alternatively, this difference could be caused by unique meteorological conditions at this site. However, when investigated further, the lower ozone values of D3 compared to D7 also appears more pronounced on weekdays (Fig. S7) reinforcing the hypothesis of industrial activities causing such differences.

U-Pod DA was the farthest away from the other monitors (~7.5km from any other U-Pod, in the north east), while DC and DB were closer together (3 km). However, it was DA and DB that have a similar spread of data around the 1:1 line, and a similar curve of data points below the main data cloud. In other words, DA and DB were more similar than DC and DB even though these two U-Pods were closer together. A possible explanation for this may be proximity to roads; DC is closest (0.6 km) to highway 91, a major freeway. Another explanation could be the environment these pods are in. DB and DA are in areas with industrial activity, whereas DC is in a more residential location.

Temporal variation in ozone values can be visually examined in more detail by singling out certain hours of data, compared to the full set. Figures 11 and 12 demonstrate this concept.

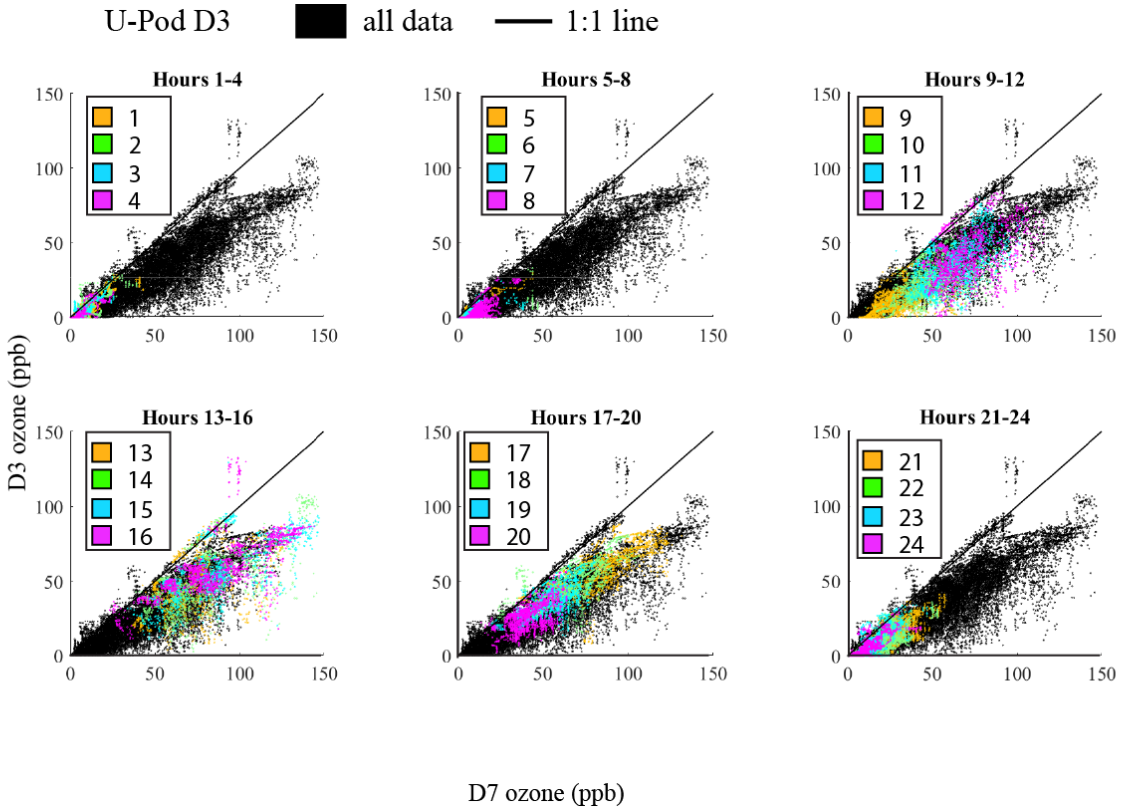

Figure 11. Data from D3, at Industrial Zone 1, plotted against D7 (at Rubidoux). In each scatterplot, colored data in the legend represents four hours of the day, and the black data represents the complete deployment dataset (all hours). The black line is a 1:1 line, not a line of best fit.

Figure 10 and 11 show that the deployment data for D3 is consistently lower when compared to D7 than the other U-Pods. D3 is 7 km from D7, in the north of the project area. U-Pod D3 was sited at a company in an industrial area where there are potentially more VOCs in the air. This site was half a kilometer from the Van Buren roadway and as such there is also the potential for elevated levels of NO_x. The NO_x reduction hypothesis posits that depending on the ratio of NO_x to VOCs in an area, increasing NO_x can increase or decrease the concentration of ozone. The titration of ozone with NO_x can deplete concentrations of ozone. The proximity of D3 to Van Buren and the potential for increased local industrial sources of VOCs affecting the ratio, may cause ozone at D3 to appear lower when compared to that of D7. Beginning in hour 9:00 and extending through hour 12:00, there were general increases in the ozone concentrations recorded, and the points start to spread out, demonstrating significant spatial variations that are temporally relevant. From hours 13:00 to 16:00, there was less of a trend in terms of generally increasing or decreasing, and values cover a large range of ozone. From 17:00 – 20:00, we observed a reversal of the trend in the 9:00 – 12:00 hour block as ozone starts to decrease again and becomes more densely clustered. The reversed color trend from left to right in these two subplots is very clear. Lastly, for the remaining hours of the day, the measurements become very dense and values decrease again, completing a daily cycle.

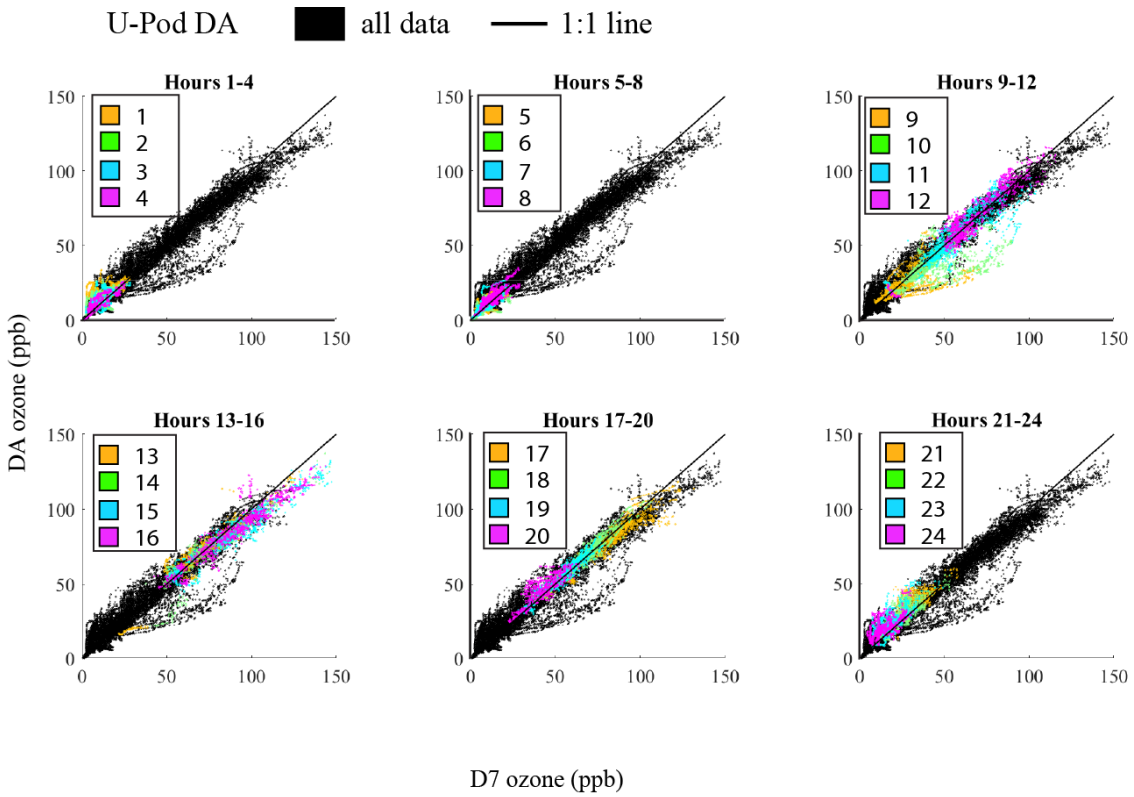

Figure 12. Data from DA, located at Commercial Zone 1, plotted against D7 (Rubidoux). Each scatterplot is four hours of the day, with the black data representing the complete deployment dataset (all hours) and data points recorded within each hour bin are marked by the colors and times in the legend. The black line is a 1:1 line, not a line of best fit.

Figure 12 shows the relationship between DA and D7 at varying hours during the day, highlighting some interesting observations. First, there was far less spread around the 1:1 for DA (compared to D3) indicating that ozone measurements from D7 and DA were more similar than D7 and D3. DA is similarly distanced from D7 as D3, about 7.5 km away, but still in the northern area of the study. These plots show concentrations from DA are more similar to D7 than those of D3, because there is much less deviation from the 1:1 line in data points. Also of interest is the strange claw shape on the underside of the black data cloud. The analysis in

Fig. 12 was conducted for all pods, but not all are shown here. It appears that many of these points occur mostly in hours 9:00 through 11:00 for all affected U-Pods. The data points from the claws in DA occur in a few consecutive hours on three different days, similar to D7. The claw in D7 is not causing this effect in DA, because they occur at different times. One possible explanation for this may be the presence of one or more gas species that is not captured by the model which affects either the sensor directly, or the concentration of ozone in the vicinity for a short time. These gases could be localized ozone precursor emissions such as NOx or reactive organic gases (ROGs) which happen to correlate with morning rush hour. This claw-shape occurs at the D0, DB, and DE sites as well, all of which are closest to Van Buren Blvd. Also, the data within this claw shape appear to happen more often on the weekend than on weekdays (Fig. S7). We do not have sufficient data on NO_x concentrations or high-resolution traffic information to draw specific conclusions about how these may be affecting ozone at different sites. This could be an area for future research.

4 Conclusions

In the region of Riverside, CA, we were able to observe spatial and temporal variability of ozone across an area of roughly 200 km^2. Field validation of sensor O_3 measurements to minute resolution reference observations resulted in R^2 and RMSE of 0.95-0.97 and 4.4-7.2 ppbv. The Thermo Scientific Model 49i Ozone Analyzer that SCAQMD uses for FRM has an acceptable measurement noise of 5% of the precision gas input, or around 5 ppbv for ozone. The measurements from the MiCS 2611 ozone sensor should not be thought of as a way to replace regulatory AQMS or prevent future stations from being built, but rather supplement that information. After all, these sensors not only depend on reference grade measurements but also the quality control and assurance carried out at those stations. These low-cost sensors can help in deciding where future AQMS be erected as well as inform the existing gaps between stations.

Technological difficulties of obtaining sensor data through environmental extremes, increased sensor variability with high ozone values, electrical issues and data retrieval are all issues encountered when using a U-Pod sensor network. Although the sensors themselves are low-cost, the data retrieval, validation and analysis are not. Data were retrieved on a biweekly basis which required a field visit to each site. Sensor platforms that wirelessly transmit data (or stream data) require additional hardware and may limit sensor placement yet are promising for many applications. The U-Pod has since evolved to incorporate wireless data transmission in some units. Processing (e.g., QAQC, filtering) and analysis of these data (~2 MB/pod/day) constitutes the majority of time for such campaigns. Future projects may involve very large numbers of sensors, therefore time expenditure for this network method needs to be reduced.

The highest amount of variability between U-Pods based on the R^2 values of all their possible pairs to occur between 21:00 and 3:00, as well as at 9:00. U-Pods are more correlated around 5:00 and the period between 11:00 and 19:00. Based on the median absolute differences between all possible U-Pod pairs, the U-Pods are most similar at 6:00, and peaks in differences (least similar) occur at 10:00 and 15:00-6:00. The uncertainty of these measurements, as determined by the validation results of D0 and D7 is 4.4 – 5.9 ppbv.

For future sensor research, an analysis of the amount of time spent collocating (calibrating) to the amount of time deployed (applying calibration) would be very beneficial for the sensor community. This information can inform how long sensors can be deployed in given region under given environmental conditions before recalibration is warranted. In this study, for nearly three

weeks of collocation time, sensors were deployed for more than nine weeks with only slightly variation of performance from week one to week nine. It is important to collocate the sensors as frequently as possible while balancing other resources. Sensor quantification using different mathematical approaches to linear regression could improve the performance. Since higher values of ozone are of the greatest interest to regulators and the public from a human health standpoint, and the sensor variability increases at those higher values, perhaps the regression could be fit differently to suit those needs. An example could be to fit a piecewise function, to better capture the low-ozone and high-ozone regimes separately, or other non-linear models.

Additionally, including measurements of other compounds in the study could help to explain causes for spatial and temporal variability in both ozone. For example, including information on nitrogen oxides could help inform the effects on traffic on these compounds, while land use data could reveal the effect of vegetation or industrial operations on measurements. Furthermore, this study was conducted in an area with relatively high levels of ozone, which can be simpler to detect. Many people live in areas that have ozone levels closer to EPA required levels, though they still experience some periods of non-attainment. To make this research more relevant to all people, the next step could be to try and detect the same spatial and temporal variability in these places as well.

Code and Data availability

The final, filtered dataset and the codes used to make the plots in this manuscript are available on Mendeley at DOI: 10.17632/j36zwxy8v4.1. All codes used to perform the linear regression are not included. Raw data are not included because they cannot be interpreted in concentrations without the regression model codes, and results from raw voltages could be misleading. Reference data provided by SCAQMD did not undergo usual procedures of quality assurance and quality control before they were provided to us.

Author Contributions

K. Sadighi helped conduct the field experiment and analyze deployment data, and prepared the manuscript with contribution from all authors. E. Coffey was the lead field scientist, performed the calibrations and validation analysis, and conducted the literature review. A. Polidori and B. Feenstra facilitated collaboration between the Hannigan group and the South Coast Air Quality Management District and provided useful information on air quality conditions in Riverside County. Q. Lv, D. K. Henze, and M. Hannigan provided guidance and academic support for the project.

Competing Interests

The authors declare that they have no conflict of interest.

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant no. 1442971 for the CyberSEES project. Thank you to members of the Hannigan group for your help: Ashley Collier, Ricardo Piedrahita, and Joanna Casey, and Drew Meyers for your support. Also, thanks to our field technician, Brandon Wong. We also acknowledge the individual participants that accommodated U-Pods during the deployment.

References

2015 Traffic Volumes. Caltrans. Available from: http://www.dot.ca.gov/trafficops/census/volumes2015/ (Accessed 3 Feb. 2016), n.d.

Annual Monitoring Network Report for Twenty-three Districts in California. Rep. California Air Resources Board, July 2013. Available from: https://www.arb.ca.gov/aqd/amnr/amnr2013.pdf, (Accessed 1 Feb. 2017).

Barsan, N. and Weimar, U.: Conduction model of metal oxide gas sensors, Journal of Electroceramics, 7(3), 143–167, 2001.

Bart, M., Williams, D. E., Ainslie, B., McKendry, I., Salmond, J., Grange, S. K., Alavi-Shoshtari, M., Steyn, D. and Henshaw, G. S.: High Density Ozone Monitoring Using Gas Sensitive Semi-Conductor Sensors in the Lower Fraser Valley, British Columbia, Environ. Sci. Technol., 48(7), 3970–3977, doi:10.1021/es404610t, 2014.

Blanchard, C. L., Tanenbaum, S. and Hidy, G. M.: Spatial and temporal variability of air pollution in Birmingham, Alabama, Atmospheric Environment, 89, 382–391, doi:10.1016/j.atmosenv.2014.01.006, 2014.

Clements, A. L., Griswold, W. G., Rs, A., Johnston, J. E., Herting, M. M., Thorson, J., Collier-Oxandale, A. and Hannigan, M.: Low-Cost Air Quality Monitoring Tools: From Research to Practice (A Workshop Summary), Sensors, 17(11), 2478, doi:10.3390/s17112478, 2017.

E2V MiCS-2611 Ozone Sensor datasheet: http://www.mymectronic.com/datasheet/13059_4173143_mics-2611.pdf, last access: 8 Nov 2017.

Environmental Protection Agency. Ambient Air Quality Surveillance (40 CFR part 58), 2006. Available from: https://www.ecfr.gov/cgi-bin/retrieveECFR?n=40y6.0.1.1.6, (Accessed 22 November 2016).

Environmental Protection Agency. National Ambient Air Quality Standards (40 CFR part 50), 2013. Available from: https://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title40/40cfr50_main_02.tpl, (Accessed 2015).

Environmental Protection Agency."Green Book 8-Hour Ozone (2008) Area Information.", 2016. Available from: https://www.epa.gov/green-book/green-book-8-hour-ozone-2008-area-information, (Accessed 21 Sept. 2016)

Gao, H. O.: Day of week effects on diurnal ozone/NOx cycles and transportation emissions in Southern California, Transportation Research Part D: Transport and Environment, 12(4), 292–305, doi:10.1016/j.trd.2007.03.004, 2007.

Jacob, D. J.: Heterogeneous chemistry and tropospheric ozone, Atmospheric Environment, 34(12–14), 2131–2159, doi:10.1016/S1352-2310(99)00462-8, 2000.

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S. and Buckley, K.: Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, Atmos. Meas. Tech., 9(11), 5281–5292, doi:10.5194/amt-9-5281-2016, 2016.

Korotcenkov, G.: Metal oxides for solid-state gas sensors: What determines our choice? Materials Science and Engineering: B, 139(1), 1–23, doi:10.1016/j.mseb.2007.01.044, 2007.

Kumar, P., Morawska, L., Mahtani, C., Basks, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L. and Ritter, R.: The rise of low-cost sensing for managing air pollution in cities, Environment International, 75, 199–205, doi:10.1016/j.envint.2014.11.019, 2015.

Lin, C., Gillespie, J., Schuder, M. D., Duberstein, W., Beverland, I. J. and Heal, M. R.: Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide, Atmospheric Environment, 100, 111–116, doi:10.1016/j.atmosenv.2014.11.002, 2015.

Littman, F. E. and Magill, P. L.: Some Unique Aspects of Air Pollution in Los Angeles, Air Repair, 3(1), 29–34, doi:10.1080/00966665.1953.10467586, 1953.

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

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M. and Fishbain, B.: On the feasibility of measuring urban air pollution by wireless distributed sensor networks, Science of The Total Environment, 502, 537–547, doi:10.1016/j.scitotenv.2014.09.059, 2015.

Peterson, P. J. D., Aujla, A., Grant, K. H., Brundle, A. G., Thompson, M. R., Vande Hey, J. and Leigh, R. J.: Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments, Sensors (Basel), 17(7), doi:10.3390/s17071653, 2017.

Piedrahita, R., Xiang, Y., Masson, N., Ortega, J., Collier, A., Jiang, Y., Li, K., Dick, R. P., Lv, Q., Hannigan, M. and Shang, L.: The next generation of low-cost personal air quality sensors for quantitative exposure monitoring, Atmospheric Measurement Techniques, 7(10), 3325–3336, doi:10.5194/amt-7-3325-2014, 2014.

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A. and Rickerby, D.: End-user perspective of low-cost sensors for outdoor air pollution monitoring, Science of The Total Environment, 607–608(Supplement C), 691–705, doi:10.1016/j.scitotenv.2017.06.266, 2017.

Simon, H., Wells, B., Baker, K. R. and Hubbell, B.: Assessing Temporal and Spatial Patterns of Observed and Predicted Ozone in Multiple Urban Areas, Environ. Health Perspect., 124(9), 1443–1452, doi:10.1289/EHP190, 2016.

South Coast Air Quality Management District (SCAQMD). Air Quality Monitoring Network Plan. Site Plan Rubidoux. Available from: http://www.aqmd.gov/docs/default-source/clean-air-plans/air-quality-monitoring-network-plan/aaqmnp-rubidoux.phttp://www.atmos-chem-phys-discuss.net/acp-2016-850/acp-2016-850.pdf.df?sfvrsn=11, (Accessed 12 Dec. 2015).

Sun, Y.-F., Liu, S.-B., Meng, F.-L., Liu, J.-Y., Jin, Z., Kong, L.-T. and Liu, J.-H.: Metal Oxide Nanostructures and Their Gas Sensing Properties: A Review, Sensors, 12(3), 2610–2631, doi:10.3390/s120302610, 2012.

U.S. EPA. 2013 Final Report: Integrated Science Assessment of Ozone and Related Photochemical Oxidants. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-10/076F, 2013.

Williams, D. E., Henshaw, G. S., Bart, M., Laing, G., Wagner, J., Naisbitt, S. and Salmond, J. A.: Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network, Meas. Sci. Technol., 24(6), 065803, doi:10.1088/0957-0233/24/6/065803, 2013.

Tables

Table 1: Field calibration results of the Linear 4T model for ozone sensors showing R2 and RMSE with the reference monitor data. Two O3 entries means there are two different sensors in the same U-Pod.

U-Pod ID	D0	D3	D4	D5	D6	D7	D8	DA	DB	DC	DD	DE	DF
R^2, RMSE	0.98, 3.1	0.98, 3.0	0.98, 2.6	0.99, 2.7	0.98, 3.5	0.98, 2.8	0.98, 3.0	0.97, 3.9	0.98, 2.8	0.99, 2.6	0.99, 1.8	0.97, 3.4	0.98, 3.1
	0.98, 3.2	0.98, 3.0	0.98, 2.7	0.98, 3.0	0.99, 2.4	0.98, 3.0		0.97, 3.9		0.98, 2.7	0.99, 1.8	0.98, 2.9	0.98, 3.0

Table 2. Overall validation sensitivity results showing mean residuals, median residuals, R^2 and RMSE of sensor measurements against Rubidoux or Mira Loma AQMS O_3 (ppbv) observations. Two-hundred iterations of 10% randomly selected minute-data were used for validation statistics (± 1 SD).

U-Pod ID	mean residual	median residual	mean R^2	mean RMSE	validation method
D7 O_3 Sensor 1	2.4 ±0.1	1.2 ±0.1	0.965 ±0.001	5.6 ±0.1	Same location
D7 O_3 Sensor 2	2.8 ±0.1	1.5 ±0.1	0.963 ±0.001	5.9 ±0.1	Same location
D0 O_3 Sensor 1	0.7 ±0.1	0.8 ±0.1	0.974 ±0.001	4.4 ±0.1	Different location
D0 O_3 Sensor 2	1.1 ±0.1	1.0 ±0.1	0.971 ±0.001	4.9 ±0.1	Different location
*D5 O_3 Sensor 1	5.5 ±0.1	5.1 ±0.1	0.971 ±0.001	5.0 ±0.1	Different location
*D5 O_3 Sensor 2	6.4 ±0.1	3.9 ±0.1	0.953 ±0.001	7.2 ±0.1	Different location

*D5 experienced an electrical issue resulting in data omission from analysis