amt-2017-183 Author Response to Reviewer Comments

Anonymous Referee #1
Received and published: 26 September 2017
The manuscript describes a system of two low-cost sensors, one for CO2 and one for O3, that were calibrated and deployed in Riverside, CA. This area of research is very active right now and this paper is timely as there are not a lot of studies that are published on low-cost CO2 sensors specifically that rigorously evaluate their performance, although many groups are working on this type of sensor. The paper is mostly well-written and organized, although there were a few confusing points to me over the language and relationship between the calibration, collocation, validation, and deployment periods as described. Some of this confusion prevents the reader from really understanding the uncertainty evaluation made for these sensors. Specifically, for the CO2 measurements, there is no clear explanation of large differences between a low-cost sensor that is co-located with a reference sensor over all the time periods. I am not convinced that the stated uncertainty of 15 ppm is valid. The authors also note some issues with large shifts in calibrations of the CO2 sensors due to manufacturer's software, leading to eliminating some of the sensors. Given these issues, I am not sure the CO2 sensor portion of this paper is useful to a reader without further clarification and perhaps additional analysis of the existing data set. The ozone sensors seem to have been more thoroughly evaluated. More on these and other comments are below. I recommend publication only after some major changes to the manuscript addressing the issues.

**The availability of the CO_2 data, due to technical issues with our certain set of sensors, has made evaluation of that dataset difficult. After further consideration, we have decided to remove the CO_2 sensor results from the paper. We feel that it makes the analysis of the ozone sensors less strong. Throughout the paper, we have added sentences that hopefully clarify for the reader which time periods are being discussed, and where the U-Pods are during those time periods. For addressing the specific comments, most or all of the reviewer's suggestions have been incorporated into the text, and these comments are marked with an asterisk indent below.*

Specific Comments:
For all the questions posed below, I would recommend the authors address an answer in the text of the paper itself, and not just answer in the reviewer responses, unless they have reason not to include the information in the text.

Introduction:
Page 2 L11: The authors state that AQMS is expensive, but that is of course a relative term - can they estimate a rough cost? Is the instrument itself the source of expense (or for example, is it the cost of maintenance, data retrieval, site access, calibration)? And in general (this may come up elsewhere), I found that sometimes it is not clear in the paper whether the authors are referring to the ozone or co2 sensors. I would think that the AQMS only monitor ozone, so this should be specified here. In general, there is a feeling while reading this paper that the initial focus was on ozone and co2 was thrown in later, so a clean read-through to look for this might be good.

* *We realized that the analysis and further synthesis of CO_2 data was not as strong as the ozone analysis and discussion. This is in part a product of the poor ELT CO_2 sensor characterization due to*

manufacturer algorithms. We have decided to remove the CO_2 analysis and instead focus this paper on ozone variability. In regards to the monitoring station costs, entire monitoring stations can cost more than 100 thousand USD in addition to the expense to run and maintain them as well as their physical footprint, usually requiring property and indoor conditioning. These details were added for better reader context. More details surrounding the time and cost requirements of the sensors were added as well.

Page 3 L4: The minimum number of sites required by whom? Sites for ozone measurements presumably, not co2? I found this paragraph confusing - are 20 sites not enough to capture the variability in concentrations that is spatially heterogeneous below 10s of km? Are the 20 sites the same as the "current EPA monitoring networks"?
**The siting requirements are stipulated by the US EPA for ozone monitoring. Many studies before ours, as cited in the manuscript have found variability below the spacing of air quality monitors 10 km apart. Also, although there are 20 sites, they are not spaced evenly throughout the county.*

In the references for other low-cost sensor experiments, there have been studies using low-cost CO2 sensors: Shusterman et al. (ACP: https://www.atmoschem-phys.net/16/13449/2016/acp-16-13449-2016.pdf) and Martin et al. (AMT: https://www.atmos-meas-tech.net/10/2383/2017/amt-10-2383-2017.pdf). These should be mentioned.
**These are useful resources, thank you for them. Analysis of CO_2 has been omitted to further investigate at a later date and these sources will certainly come in handy.*

P3 L31: should be "high VOC concentrations". (and presumably NOx?).
* *Yes, thanks, this is what was meant. Fixed to include "concentrations" and "NOx".*

P4 L13 "should be there is a large number of vehicles"...
Suggested modification made in text.

P5 L8: Some indication of why medians were used rather than means? (to reduce influence of extreme outliers?).
** Median values were used to reduce the influence of outliers within the minute. This clarification has been added to the text.*

P6 L7: semicolon I think should be a colon.
* *This change was made.*

P6 L10: It seems that uncertainties and precision should be given for the two reference sensors. For the CO2 standards, who certified them and what is their associated uncertainty? Is the Licor calibrated or drift-corrected in the field at all? What about the ozone sensor - how is it calibrated or drift-corrected? It would be important to assess whether either of these instruments is sensitive to ambient temperature, pressure, humidity, etc., same as the low-cost sensors. If these are not corrected or controlled for these variables, the authors should address whether this fact changes their interpretation of the various correlations and fits, and how. i.e. if the ozone reference concentration is dependent on temperature, would that have resulted in the interaction term that was observed? (I'm not sure, perhaps not.).
* *The uncertainty of the reference ozone monitor is discussed later in the paper during the discussion and is framed within the uncertainty of the sensors. We have added more information regarding temperature and humidity compensation for the reference ozone monitors.*
Analysis of CO_2 has been omitted to further investigate at a later date.

P6 L18 sentence structure awkward
* *Thanks for catching this. Modified sentence for consistency*

P6 Eq 2 p1 should have the 1 subscripted.
* *Yes, fixed in text.*

P6 bottom, P7 L1: Martin et al. did this for CO2 sensors, but only for a much shorter period of time (2 weeks?).
**Thank you for turning us to this. Analysis of CO_2 has been omitted to further investigate at a later date.*

P6 L30 perhaps to add clarification here, note that the comparisons were made between the concentrations from the low-cost sensors after being corrected by the equations 1 and 2 using the coefficients from the initial calibration test (constant coefficients p in time?) and the reference concentrations.
* *To add clarification to this sentence, the first two sentences of this paragraph were combined. Now it is clear that raw sensor data was converted to meaningful gas concentrations using the calibration model coefficients and those values were compared to reference-grade concentration measurements over the ~3-month validation period.*

P7 L5: microenvironmental space?
** Yes, this refers to the small environmental characteristics of a location (e.g., temperature, humidity, pressure, gas species concentrations etc.) but we have removed this word for clarity.*

P7 L12: this is the first reference of the 4T equation - this is equation 1?
* *Yes, this is Eq. 1. Clarified in text.*

Table 1: I wonder how different the coefficients Px are between different individual sensors? Can the authors give an idea of this?
* *This is an interesting question and has prompted substantial modification to this manuscript. Although this wasn't initially within the focus of the paper, we thought it would be appropriate to add now that the CO_2 analysis will not be included at this point. A figure showing distributions of modeled coefficient values and coefficients of variation for the sensors has been added along with a short discussion on similarity among sensors and potential reasons for these differences.*

P7 last line: "during the deployment that were outside the range of those experienced during the calibration time period" might be clearer here.
* *Yes, this would help. This has been clarified in the text*

P8 L2: "As such" is not really clear as to which path you chose (assess or avoid). So you went with avoiding any extrapolation and filtering any data points with parameters outside the calibration range?
** We did not want to extrapolate in order to reduce error. The second sentence here is correct. This was further explained in the text.*

P8 L15-20. Not clear - wouldn't it be best to just eliminate O3 values that were higher than those experienced during the calibration period? How could measurements be over 7% of the highest maximum value, if it was the highest maximum value? Is it because you are looking at the highest value of the reference instrument? This seems

odd to me all around. Why filter O3 and not Co2?

* *We did not want to eliminate values that were higher than seen in the calibration because high ozone is of special importance to human health, and those values occur later in the summer. This is one of the reasons we filtered for environmental conditions, to reduce some of the error while still being able to record high values of ozone. In addition, the validation period acts as a way of knowing how well we estimate those higher ozone levels during the deployment period. We have changed the way that we filtered for maximum values, which has been explained in the text, and now makes the maximum 171 ppb. We feel this method is appropriate and did not significantly change results. A total of 110 data points were influenced by this change across all the U-Pods.*

P9 L 13: So "calibration validation" refers to the deployment period from the previous section? Would be good to clarify that the validation period is the same as the deployment period, since both words are used here. This paragraph makes it sound like the best model was chosen based on how well each model did during the validation period, not using the same coefficients and model necessarily that were chosen during the calibration? The authors should clarify - I would have thought that the model was built using the calibration data set (including specific coefficients) and then applied during the deployment/validation, and then that corrected data would be compared with the reference sensor. Can the authors confirm that the coefficients from the calibration period were used in the validation period, and that the validation period is the same period as the deployment period?

* *Thanks for pointing out this confusion. Figure 1 has been updated to make this clearer. Validation refers to U-Pods that were still located with a reference station during their deployment. The coefficients were arrived through the collocation with the reference station during the calibration time period. We tried to fit several different models to the calibration period data (Table S1). Then we took those coefficients (from multiple models) and applied them to data during the validation time period, which is the beginning of the deployment. The model that performed the best on the validation data was applied to all the raw data for both calibration and deployment.*

P9 L24: What does a precision check entail exactly? This should be stated in the text.

* *The details of the precision check were added to the text.*

P9 L25: wording: should say "5% from expected values (corresponding to a concentration of about 5 ppb), subsequent data would be flagged ...".

* *Agreed. Change made.*

P9 L26 awkward again: "Values within 5 ppb of the expected value would not be flagged".

Changed this sentence to be more clear.

P9 L31: How large was the bias on this D45 UPod, and was this included in the statistics given above for ranges for the mean and median residuals? L34: This is confusing, as statistics were already given above. Is this 1-2 ppb bias based on mean or median residual?

* *The mean bias residual for D5 was 5 - 6.4 ppb. This may have been confusing due to the placement of Table 2. We have moved it ahead of this paragraph and clarified the types of statistics. Because D5 was later omitted from the analysis due to the electrical modification, these statistics were not included in the overall uncertainty of the sensors. However, we made sure to show that bias to be transparent about potential issues of using these sensors.*

P10 L1-2: Only one CO2 sensor was co-located with a reference for CO2 during the

validation period? Maybe this can be re-stated here for those of us who got confused as to why only one sensor was used to assess this uncertainty.

** Analysis of CO_2 has been omitted to further investigate at a later date.*

P10 L1-2: were these higher concentrations and higher humidity values within the range observed in the calibration period, or where they extrapolations of those fits?

** These were not extrapolations, as explained in comments above. Filtering for the environmental variable space had taken place.*

Figure S6: These are plots made for the validation period, not the calibration period, so the fits shown as lines are not used to correct the data, just for informational visuals, is that right? is the red line in (a) the 1:1 line or the linear fit? (same comment for S5).

**Yes, the red lines are to show where the 1:1 is for viewing purposes. Changes were made to the captions of those figures to make this clear.*

Table 2: for CO2, the RMSE is of the 1-minute data, while the mean residual over the whole period (how long was this period again and during what season?), is much lower at 3 ppm. What would the RMSE be for 1-hour means? Later in the paper hourly means or medians are used to look at differences/trends/etc., so this is the more relevant metric. If averaging the sensor data even further to 1-hour averages comparing the 1-hour medians reduces the RMSE that would be useful to know.

** Analysis of CO_2 has been omitted to further investigate at a later date.*

P11: Deployment: Is this the same as the validation period? Also in this first paragraph the collocation period is referred to - please confirm and state clearly that this is what was used earlier as the "calibration" period, i.e. the period when all the sensors were collocated and the coefficients and models derived.

** This has been addressed in previous comments. Validation just refers to the U-Pods that were deployed to places where there were reference monitors. We added a sentence at the beginning to clarify, as well as at the beginning of 3.3.*

P11 L9-10 - I agree on the usefulness of comparing variability during calibration period vs. during deployment - except for the additional uncertainty caused by calibration drift over time, which cannot be assessed with the current data. This should be noted as a caveat - the true uncertainty during deployment might be larger because of the drift in the coefficients and model that is used. This is one of the key questions about use of low-cost sensors in the field - how often do they need to be re-assessed or calibrated?

** As mentioned earlier this is an important topic for sensors. It should be noted that the calibration model we used incorporates time so re-calibrations were not done over the deployment period. Rather, the validation serves as a reference for how well the pods that weren't collocated with reference monitors during the deployment are performing. We added more analysis of model performance over time (via the results summarized in Figure 7) and discussed possible reasons for drift.*

P11 lines 16-20 - please mention the time period, time of year of these measurements. Also this section is a bit repetitive with the next paragraph on P12, lines 6-10, which states the same information about how we would expect the diurnal cycle to look. Perhaps merge?

** We have included more details about the time periods that Gao used to show the diurnal cycle of ozone, and excluded a repetitive sentence from P12, lines 6-10.*

P12 L13 and later in the text, when examining pair-wise R^2 values, are the pairs of

sensors that are in the same location excluded, so that we are only evaluating sensors that are in different locations? My understanding from earlier in the text was that there were 2 ozone sensors in each location even during the "deployment" period. [I am now re-reading the earlier text and realize that there were two ozone sensors in many of the U-Pods - in this case, which sensor's data is being used?]. But still, during the validation/deployment period, some sensors (D0 and D5) were at the same spot - are they shown in blue in Figure 5, rather than part of the red boxes?

* *Because we excluded U-Pods D4, D5, D6, D8, DD, and DF, each of the remaining sites had only 1 U-Pod each. A sentence of this description was added to section 3.2. Blue boxes contain only data from the calibration period, when all the U-Pods were together. Red boxes have only data from the deployment, when no U-Pod is collocated with another. Also, we included a brief description of how we decided which sensor to use in the analysis in that same section.*

P12 L14: "The larger the spread and magnitude of the Rˆ2 values, the more spatial variability...". This seems backwards - the lower the Rˆ2, the more spatial variability there is. Reword?

**Correct, larger spread and smaller magnitude is indicative of more variability. This has been made clearer in the text.*

Figure 5 caption ends with "U-Pod"?

* *An unfortunate mistake, this has been removed.*

P13 L6: "The U-Pods are more correlated" should be "the U-Pod O3 measurements, after the correction using the LT4 model, are more correlated ..."

* *This change adds to the clarity; it has been incorporated.*

P13 L15: "between pairs of U-Pods".

* *Changed.*

Figure 6: X-axis should indicate (here and elsewhere) that this is local time.

**Added a sentence under figure 4 to clarify this.*

P14 - the description jumps around in time a bit here (discussion of morning, then 15-17, then back to morning again...). This is a nice analysis and necessary to accompany the Rˆ2 analysis - two values might correlate within an hour, but that could be simply because O3 is increasing across the whole basin during that hour because of PBL changes, whereas the absolute differences indicate real spatial variability in the signal.

**This paragraph is organized by looking at the inverse relationship between R^2 and differences, instead of by time of day. A sentence was added to try and guide the reader. More in-depth synthesis of the time-of-day effects are discussed later.*

P15, Lines 14-15: Was the U-Pod O3 measurement at D7 calibrated against the reference sensor during the deployment phase? i.e. the models and coefficients were re-calculated for the second phase as well? This sensor could give an idea of my earlier question which would tell us how well the calibration does over time. I.e. you could correct the U-Pod data using the calibration from the calibration period, and apply it to the deployment period, and then look at the errors relative to the reference. Re-reading table 2, it seems this is exactly what Table 2 is showing. Was there a trend in the error between the reference and U-Pod measurements over time during the validation collocated phase?

**Again, it appears that this topic was not explained well enough, so text was added to make this clear. The sensor calibration model for ozone (4T) incorporates time within the model itself so the intent of the validation using D7 was to show how well this model performed over a range of temperature, humidity, ozone concentration and time for the entirety of the deployment period. Model coefficients were not "regenerated" during the validation (deployment period for D7) as this would change the meaning of the validation statistics in the context of the pods that were not collocated with a reference monitor. Table 2 shows the distributions of the validation statistics for a random 10% of the validation minute-level data iteratively selected.*

P15 L16: "as well as hourly trends by pod (Fig. 8)." Changed.
figure 8 (& Fig 9) caption says "Each scatterplot is four hours of the day". But it's not - each plot shows all hours of the day in black.

**We have re-worded this caption to be more clear about what data is represented in the plot.*

P18: would a time series plot of these hours help interpret this weekend feature?

**A time series plot did not reveal any interesting features, but residuals helped to narrow down their time frame. Figure S4 was an investigation into what temperature and humidity corresponded to those data points.*

P19: L16: "carbon dioxide distributions" should be "CO2 distribution".

** Analysis of CO_2 has been omitted to further investigate at a later date.*

Figure 10: Looking at this figure, knowing that D7 is in the same spot as Rubi, I wonder how they compare. (similar to my previous comment). I realize that the minute data showed a 15 ppm RMSE, which seems consistent with Figure 10 left panel.
In Table 2, this 3.0 ppm median residual is for the validation period - but in Figure 10 it seems to be for the Calibration period.
But on the right panel, it would be nice somehow to do an evaluation of the D7 sensor during the deployment, using the calibration from the calibration period. The bias here seems much larger, with the median of the D7 sensor significantly higher than the Rubi Licor. Perhaps this is one of the cases of the large baseline shift? (p19 L 18 says that D7 observed higher carbon dioxide - but the Licor did not observe that same high level, so this is not an accurate way to characterize what appears to be sensor drift).
P20, L1. The uncertainty of 15 ppm for CO2 was determined as an RMSE of 1-minute data, where the median difference was only 3 ppm. 15 ppm does not seem to be the correct uncertainty on the median of a distribution of hourly data over a several-monthlong (?) period.

* *Analysis of CO_2 has been omitted to further investigate at a later date.*

Without more analysis or elimination of sensors that had large shifts in calibration, this claim does not seem to be supported - it is still not obvious that the sensors can determine spatial variability in CO2. Moreover, noting that some sensors are giving hourly mean values (albeit outliers, granted) that are close to 300 ppm (after the calibration correction!) makes me very doubtful as to their performance.

* *Analysis of CO_2 has been omitted to further investigate at a later date.*

Back to Figure 1: the time line indicates "post-deployment" - is this referred to anywhere else in the text?

** Initially we had tried to collocate all the U-Pods together again at the end of the deployment. However, by this time the U-Pods had been untouched for some time and the data was degrading. We lost*

several Temperature and RH sensors during this time, and did not get enough data to include. This time period has been removed from figure 1, and is not used in analysis.

Figure 12. Looking again at the D7 CO2 data, this diurnal plot does not seem to match the plot in Figure 11 on the right side for this sensor, whose overall median reading was _460 ppm. Is there an error here?

* *Analysis of CO_2 has been omitted to further investigate at a later date.*

P21 L10 - the time period and season should be mentioned here again, as obviously the time of year affects these diurnal cycles.

**Yes, thank you, we have added that information to the text.*

P 21 L15. Again here the 15 ppm RMSE for CO2 is on 1-minute data, when the spatial variability etc is being evaluated using hourly medians for the most part.

* *Analysis of CO_2 has been omitted to further investigate at a later date.*

P22, L1-8 - these are very good points to make here, as I think these costs are often overlooked in the context of the low-cost sensors.

**Thank you.*

P22 L15 should be "as high correlations with each other"

**Yes, the change was made.*

P22 L15-17. I do not think this has been shown here.

* *Analysis of CO_2 has been omitted to further investigate at a later date.*

P22 L19: How often is frequently? This remains to be seen in future work perhaps. Is there a way to determine this frequency from the data collected during the validation period? (i.e. is there a drift relative to the standard with time?). The authors did choose subsets of the validation data in order to do the evaluation in a more robust way, but an investigation here of the time-dependence of the errors would be useful.

* *The question of how often sensors need to be recalibrated is indeed an important and widely discussed topic. Suggested durations between calibrations are highly dependent on the environment and gas species of interest. Due to the interest in this topic, an analysis and paragraph presenting the results of validation from the first week (directly after the calibration) and the last week independently. Residuals through time are discussed to inform the frequency of calibrations pointing to potential sources of this drift.*

P23 L3: "undergone" should be "undergo".

**Fixed this, thanks.*

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt

Reviewer 2:

This manuscript describes the deployment of low-cost ozone and CO2 monitors in Riverside,

CA. Many groups are working on these sorts of low-cost sensors, but to date most papers have focused on calibration and sensor characterization. As the authors note in the introduction, few papers outline deployment of sensors to quantify variations in air quality. Therefore, this manuscript's description of a deployment of ~10 sensors represents a meaningful contribution to the field.

**Thank you for your input on this manuscript. After more consideration, we have decided to remove the CO_2 analysis portion from the paper, and to make the ozone analysis stronger. As such, many of the comments addressing CO_2 have not been answered completely, as those sections are no longer included. The specific comments laid out below have been addressed with an asterisk here, as well as in the text, where appropriate.*

-Section 2.2 - The field calibration needs more explanation. What are the coefficients p? Why is the functional form of the calibration for CO2 and O3 different? I am aware that this group has written previous papers on sensor calibration, but not all readers will be familiar with that work. A more thorough, stand-alone description of the calibration method is required.

** A more through stand-alone description of the sensors and the calibration technique was added for clarity. The discussion of the coefficients has been addressed from reviewer 1.*

-The comparison of calibration versus deployment performance is useful (e.g., Figures 5 and 6), as these illustrate that differences in O3 measured during the deployment period are "real". However, given that Fig 7 shows that most of the U-Pods are correlated, Figure 6 would be more effective if the raw (rather than absolute) concentration difference was shown.

**For this analysis, we wanted to determine if there was spatial variability. The absolute value told us how different the sites were. The next step could be to use the raw concentration to determine the directions of difference, but as an initial try, that method is more complex and was difficult to clearly display for the reader.*

-For cases in Fig 7 where deployment data cluster around the 1:1 line - how can authors be sure that this is true variation and not some sort of uncertainty? One might expect less scatter around the 1:1 line for training data than for deployment data. E.g., if 50% of the calibration data is randomly held out as a test set, what would these scatter plots look like for the held out portion of the co-location period?

**This question is not completely clear to us. We think you are saying that there could be more scatter in the blue data if we only used 50% of the data. A similar test to this was performed with the validation dataset. For D0 and D7, the validation dataset was iterated 200 times with 10% of the data each time to address this issue. Validation results suggest our measurement uncertainty is between 4.4 and 5.9 ppbv and scatter more than this is variability.*

-Figure 8 - I do not understand the point of this figure, and it needs to be discussed more thoroughly. My read on it is that D3 is systematically lower than D7 at all times of day.

**That is correct, that D3 is lower at all times of day. It is different than all the other U-Pods, and therefore worth discussing. We have fleshed out this discussion to explain why that might be the case and go in more depth with hour-of-the day analysis.*

-Figure 9 - Do all of the data forming the "claw shape" come from the same day? Or was this phenomenon observed across a number of days?

**They come from three groups of a few consecutive hours, similar to the claw effect observed in D7. These claws are independent of the validation claw seen for D7 in the first week of validation.*

-The discussion of CO2 needs to be better integrated into the manuscript. At times it reads like

the manuscript was written for ozone, and CO2 was an afterthought.
**Analysis of CO_2 has been omitted to further investigate at a later date.*

-Page 3, Line 4 - Where does the information on the number of monitors required for Riverside-San Bernarndino come from?
** Corrected by responding to referee 1, this information is provided by the US EPA and codified in the code of federal regulations.*

-Page 4 - Line 2 says that the study area was 314 km^2 but line 4 and Fig 1 suggest a much smaller area. Please clarify.
**Thank you for pointing this out. A circle that encompasses all the pods except DD has a radius of ~8km, about 200 km^2. This was the resultant area used for the analysis.*

-The discussion of auto-ranging for CO2, and how it was dealt with, are confusing. I cannot tell if that data was removed or somehow corrected. If corrected, what was the procedure?
** Analysis of CO_2 has been omitted to further investigate at a later date.*

-The manuscript has both an Appendix and an SI, which seems redundant. The plots in the Appendix should either be in the main text or in the SI.

* *We have removed the Appendix and moved plots to the SI.*

Anonymous Referee #3

In Sadighi et al., the authors describe the calibration and deployment of the UC Boulder U-Pod for measuring ozone and CO2 in the Riverside and San Bernardino counties for a period of three weeks. The dataset is meaningful as it discusses not just validation of the low-cost sensor but also an actual deployment.

GENERAL COMMENTS
I believe the contribution of the authors is valuable and should be published in AMT if major revisions are provided. Overall, I found the structure of the manuscript to be confusing at times to follow – the authors make specific references to various units as D0, DA, DB etc without many reminders to the significance of the deployment locations. Sensor performance should be put in the context of the expected nearby sources (e.g., it makes sense that the sensor near Highway 91 would be different). I also think the manuscript is lacking in synthesis of the findings – many of the findings are stated without much interpretation.
I also think both the calibration approach, pre- and post-treatment of the data and the calibration models should be discussed in significantly more detail. You mention the significance of a time variable in the calibration, but the magnitude and direction of this coefficient isn't discussed – would this not be critical for other uses to decide on an appropriate sensor? I am also a little concerned on how data was screened to be included in the manuscript. I don't think it's very clear what data were omitted – you make some reference to only looking at data where the deployment exactly matched the calibration range. Why did you do this? Why would you expect your model could not extrapolate? I think given the structure of the mode extrapolation should be discussed.

Lastly, I would suggest you comment on how this study might translate to other areas – Riverside has some of the highest ozone in the U.S.; I would imagine that the sensors would have the least difficulty accurately measuring ozone in this area. What about other areas of the US that have occasionally observe high ozone concentrations but where it is more erratic (e.g., Pittsburgh can be in non-attainment, but average ozone is generally « 30 ppb.) Some generalizations on the findings would strengthen the paper. Lastly, I echo other reviewers when they say the CO2 discussion as written does not seem to add much value to the paper. I am also not clear on the value of the results or the impact of this auto-calibration which is underdiscussed in the manuscript. Specific comments follow.

** After much consideration, we have decided to remove the CO_2 analysis portion from the manuscript. The availability of the CO_2 data, due to technical issues with our certain set of sensors, has made evaluation of that dataset difficult. We feel that it makes the analysis of the ozone sensors less strong. We have also added a few figures to make the ozone analysis stronger. Figure 5 shows information about the coefficients of the calibrations and Figure 7 and the surrounding discussion shed more light on the topic of drift through time. We have also moved some of the figures in the SI to the main body of the paper, to flesh out the discussion on the calibration itself. More than one reviewer was confused about the time periods discussed in the paper, which we have made an effort to clarify this everywhere.*

** The purpose of filtering the deployment ozone data by environmental characteristic was to reduce the error in those values. We didn't train the model to use values outside of the calibration range. The purpose of this paper was not to test whether the model could extrapolate, but analyze spatial differences on the best data we could. We have conceded that LA is an easier place to measure high levels of ozone, but that this study does not speak to those people living in areas with lower levels of ozone, but periods of non-attainment. In the conclusion, we propose this as an important area of future research. Some of the general comments are more thoroughly answered in addressing the specific comments below, marked with asterisks*

SPECIFIC COMMENTS

P3 Line 6-7: I am not sure why you mention the CyberSEES project or what it is. Delete?

** This research was conducted, in part, under the CyberSEES project which we mention in the introduction. We feel it is important to recognize the project under which our research is conducted.*

P3 Line 17: What is this site C?

**It is a third site in that study, clarified in the text.*

P3 Lines 13-16: In this paragraph you list several findings, but there is not much synthesis of the results or general conclusions. The narrative should be improved.

** We have added more synthesis in the discussion section of the paper and framed our results with potential explanations for such trends including proximity to roadways, weekend vs weekday activities as well as meteorological changes. We focused our analysis to show variability existed and that low cost sensors have the capacity to quantify these and less forming hypotheses on what is causing these differences which has been a well-researched topic especially in the LA area.*

P3, Line 33: Is it really an ideal test bed? Wouldn't this be the best case scenario where you a) have lots of ozone and b) and lots of sources? I feel like to truly test this question, you would need to try the sensors somewhere more representative of federal average ozone concentrations.

* *We have edited this paragraph to add clarity to this statement. This region is a good test bed for answering our research question due to the reasons we mention in the text. There is increased interest from the regional governmental regulatory body in sensors and due to the potential health consequences from ozone exposure in this area, which makes it a good candidate. Testing our research question (detect variability knowing our uncertainty) in this region is the first step and then expanding to other locations is the next step.*

P5, Section 2.1: I don't think enough is said here about the long term drift or crosssensitivity of the sensors. What is the expected life span? You should be explicit about the tradeoffs.

* *A more detailed discussion and explanation of the sensor characteristics are described in the text body here. Specific life spans are not mentioned by manufacturers, but long term drift and calibration frequency are discussed further as addressed from Reviewer 1.*

P6 Line 6: What do you mean by relationships? Linear regressions? You should be specific early on.

* *We did mean linear regressions. The term "relationships" and "regressions" have been clearly defined and identified in the text. For reference here, multivariate linear regressions (with interactions) were used.*

P6 Lines 20-28: What are the coefficients? What are their directions? Can you discuss the physical meaning behind the calibrations? How significant was the drift over three weeks (i.e., what is the size of p5?). If it is large, then are these monitors really suitable to replace EPA reference monitors? These sorts of details are critical to this paper.

**These are great questions and the sensor community is especially interested in these questions (hence previously addressed these comments from Reviewer 1). Therefore, more analysis of the coefficients and dispersion was added to shed light on how these sensors perform over time and variable space. The validation set offers the most information regarding long term accuracy and integrity. It is important to note that we are not suggesting to replace AQMS but rather employing these sensor networks to supplement AQMS as needed to answer specific research questions. Figure 5 and the surrounding discussion hopefully addresses some of these questions. The median value of p5 is quite small and close to 0. We have also included Figure 7, which shows the difference between the validation from week 1 to week 9.*

P7, Lines 10-20: Again, I think some of these model calibrations are critically important and should not be in the SI. The coefficients and their interpretation should be front and center in the paper.

* *We have moved some of those figures into the main paper and decided to focus the attention on the best performing model (4T, Eq.1) and thoroughly explained it in detail.*

P8, Lines 1-9: I am confused by what this means? Did you only use data where there was exact overlap between calibration and deployment for T and RH? i.e., you did not try to extrapolate from the model? It's not clear to me what you mean in this paragraph. Also how many RH sensors failed? A number or percentage would be helpful to assess whether deploying these low cost sensors is a feasible alternative to EPA monitoring.

* *There has been some confusion here. This comment has been addressed in responses to Reviewer 1. To add more here: Four sensors had un-recoverable data, all of them experienced at least one week of missing data, and 2 failed outright.*

P9 Line 1-2: How did you deal with the data from the two ozone sensors? Did you average? Or choose one? This should be briefly mentioned.

* *This has been clarified in the text. After finding there is very little difference in the estimates from these duplicate sensors, we used the best performing (higher R^2 and lower RMSE) sensor in the calibration period in the deployment analysis. Also, see Reviewer 1 comments for more information*

P9 Line 5-10: What is this auto-calibration setting? I think more detail is needed here about what the auto-calibration does, and how you corrected for it.

**Analysis of CO_2 has been omitted to further investigate at a later date.*

P9 Line 19: "We thought that" is informal. Why do you think that? Describe, and also use more formal scientific language.

**The language here has been adjusted in the text.*

P9 Line 26: Not sure what "Values that are less than around 5 ppb different" means.

* *Changed the language to be more clear*

P9 Table 2: I think you need to do a better job describing how to interpret the results of Table 2. I am not clear on the takeaway.

**Table 2 shows the validation statistics from the interactive technique described in the text. Uncertainty from the validation (RMSE from D0 and D7, D5 was omitted) is the uncertainty of the U-Pod measurements.*

P13 Lines 11-14: You should provide some rationale for these differences and similarities by time of day. Is the 9AM difference due to rush hour? Why not at 8AM? Please expand.

**Additional context and discussion were added to help the reader know what activities are happening around these time frames. We have commented that differences in ozone could be related to rush hour times, but since our U-Pods were not equipped with sensors to detect gasses that would indicate more or less traffic, we cannot be more specific about when these times occur.*
This comment was also addressed from Reviewer 1.

P15, Lines 1-8: So is the statement here that the disagreement between the R2 metric of spatial variability and the absolute differences is real? Does it not then follow that lower R2 between two sites is not a good predictor of spatial variability? You should be clear on the implications of the findings. From my perspective, if two sites are different based on absolute concentrations, that is the better marker of spatially variability. Especially because R2 can be disproportionately affected by high leverage points in a linear regression.

* *We expounded on these finding and their differences in the text, specifically explaining the benefits and drawbacks of each comparison technique. We agree in part, absolute differences do present a more telling indicator of spatial variability, and correlations are more susceptible to misinterpretations or leveraging from outliers.*

P16, Lines 7-10: Are you sure this is really spatial variability and not some confounding factor with the sensor? I am not convinced the data you have collected here is sufficient to claim that the interesting features of D0, DA, DB and DE are related to spatial variability.

* *This is a reasonable hesitation to have. When the scatter/dispersion of the measurements is compared to the expected uncertainty we found during the validation, we are certainty seeing variability. Now for the claw shape, we agree, and have changed the narrative to actually suggest some sort of confounder acting temporarily for a few hours at most. Please see the added analysis and discussion around Figure 7.*

P16, Line 24: Proximity to a major highway would be a critical factor in differences between two sensors, even if they are <2 km apart. There is a very near-road effect of NO+O3 forming NO2 – is the O3 at site DC consistently lower? The direction of these differences should be included.

** Details regarding the exact explanations of the higher and lower values observed are limited to hypotheses but we do mention proximity to highways and the influence of NOx and VOCs as potential reasons why some sites/measurements are lower or higher.*

P18 Line 11-12: Please rephrase the question as a statement – inserting a question like this seems gimmicky. It is up to you as the researcher to provide a hypothesis for the observation and inform the reader succinctly what you observed and whether this matches expectations.

** We have rephrased these lines.*

P18: Lines 15-17: Again, I am not convinced this isn't just an artifact of the sensor. You should be clear on the potential uncertainty.

**The uncertainty of the ozone sensors is well referenced using the validation from D0 and D7. Analysis of CO_2 has been omitted to further investigate at a later date.*

P21, Line 11: Some hypotheses should be provided. The discussion should be more than statements of observations but also include some scientific assessment of what was observed. What kind of future investigation is warranted? What would that future study look like?

** Discussion and synthesis of the observations and trends is boosted but our main goal is to show that the sensors are viable for detecting variability on the scales (temporal and spatial) mentioned. Analysis of CO_2 has been omitted to further investigate at a later date.*

P22, Line 1: This is the first time, to my knowledge, that you directly state that the MOx sensors cannot replace EPA monitors. Some space should be devoted to discussing the suitability of the U-Pods for monitoring and what is a reasonable expectation from the units.

** A clear discussion of the limits of the sensors and their desired applications as we see it has been added here and a clear message has been added to the beginning of the paper that the intent is not to replace the AQMS but rather to supplement them for more information.*

P22 Line 3-6: Can you expand on the computation time, people demand, etc to give a sense of the rigor involved in the deployment as a guide to others?

**Additional information on time resources of using the sensors was added.*

P22 Line 19: What is "frequently"?

**As addressed in Reviewer 1 and 2, additional discussion has been added on this point, including figure 7, which show the difference between the validation of week 1 and week 9.*

TECHNICAL CORRECTIONS

P1, Line 11: I am not sure it's grammatically correct to say the tool provides low-cost sensors. This just didn't make sense to me.

** We changed "provides" to "contains". The U-pod (the tool) contains sensors.*

I would replace every instance of "Author and coworkers" with "Author et al." – e.g. on P2 Line 29, and elsewhere in the manuscript.

** We have taken this suggestion and made changes in the text.*

P3 Line 4 and 5 “the Riverside-San Bernardino counties”
**Changed*

P12 Line 6: Some word missing here in “Figure 4 gives some context temporal variability”
* *This was changed.*

P22-23, Line 31/Line 1: You should just delete the sentence about it being a lot of work.
**This was removed.*

P23, Line 3: “undergo” vs “undergone”
* *This was changed.*

P24: Why is there an appendix that is separate from the Supporting Information? I would just put in the SI?
* *This has been addressed by a previous comment. It has all been combined into the SI.*

Intra-urban spatial variability of surface ozone ~~and carbon dioxide~~ 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 ~~such as~~ ozone (O_3) ~~and carbon dioxide~~. 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 ~~and CO_2~~ in a 200~~314~~ km^2 area of Riverside County near Los Angeles, California. This tool ~~provides~~ 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 ~~and CO_2~~ measurements to minute resolution reference observations resulted in R^2 ~~squared~~ and root mean squared errors (RMSE) of 0.95 – 0.97 and 4.4 – 5.9~~7.2~~ ppbv ~~for O_3 and 0.79 and 15 ppmv CO_2~~, respectively. Using the deployment data, ozone concentrations ~~and carbon dioxide concentrations~~ were observed to vary on this small spatial scale. In the analysis based on hourly binned data, the median R^2 ~~squared~~ 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 ~~ppb~~ ppbv for the ozone deployment. ~~For carbon dioxide, distributions of all measurements vary from 413 – 425 ppm during the calibration (collocation) and 406 – 472 during the deployment.~~ Since median differences between U-Pod concentrations during deployment are larger than the respective root mean square error values ~~for ozone and carbon dioxide~~, we can conclude that there is spatial variability in ~~these~~ this criteria pollutant~~s~~ 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. ~~For our CyberSEES project (NSF Award ID: 1442971), the measured variability between AQMS will provide validation for atmospheric model downscaling to the 10s of kilometer scale – smaller spacing than current EPA monitoring networks.~~

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 ~~so as~~not as ~~a~~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 ~~the~~ 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.

~~Carbon dioxide measurements are a special concern to organizations such as the California Air Resources Board (CARB) and others for providing information on combustion emissions as well as greenhouse gas concentrations (Verhulst et al., 2016). One of the most difficult aspects of measuring CO_2 in urban areas is the collocated nature of anthropogenic and biogenic sources and sinks, and difficulty in isolating these components (Verhulst et al., 2016, Hutyra et al., 2014, Kort et al., 2013). However, accurate measurements of carbon dioxide in cities will become increasingly important for carbon cycle and climate change science (Hutyra et al., 2014). This will be especially important in a Megacity like Los Angeles, where CO_2 is highly variable across the region (Verhulst et al., 2016). Kort and coworkers suggest that networks of carbon dioxide measurements with high spatial and temporal variability are necessary to properly study greenhouse gas fluxes in the Los Angeles area (Kort et al., 2013). Hutyra and coworkers also stress the importance of complete datasets that also report the level of uncertainty (Hutyra et al., 2014). Low cost sensor networks could help to tackle some of these challenges. The analysis of carbon dioxide data in this study seeks to explore the extent of spatial and temporal variability, while framing those findings in measurement uncertainty.~~

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 ~~For our CyberSEES project (NSF Award ID: 1442971), the measured variability between AQMS will provide validation for~~

~~atmospheric model downscaling to the 10s of kilometer scale – smaller spacing than current EPA monitoring networks.~~ [~~T~~he current spacing is not sufficient to capture high spatial resolution of concentration variations (Bart et al., 2014; Moltchanov et al., 2015). ~~Additionally, this~~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).

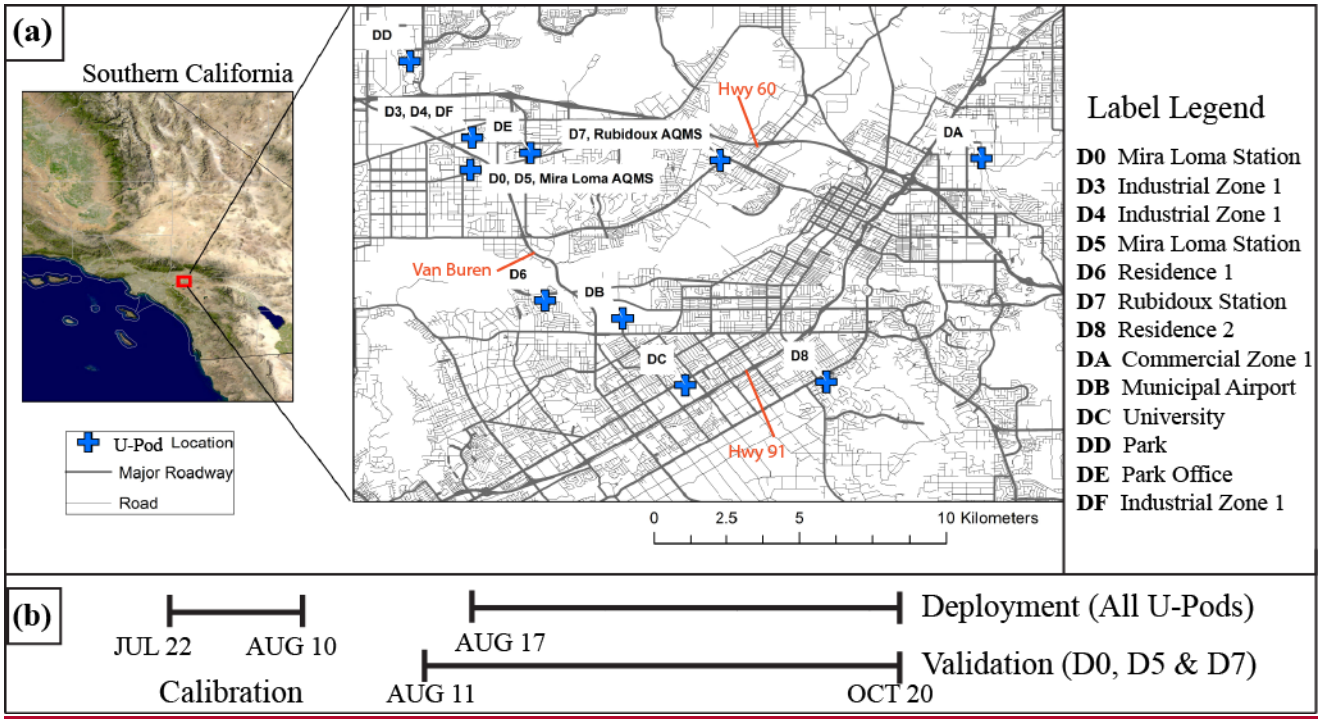
Networks of air quality sensors have been deployed in various settings. Moltchanov ~~and coworkers~~et al. measured O_3, NO_2 and VOCs in Haifa, Israel in the summer of 2013 to test the viability of sensor networks measuring small scale (100s of meters) intra-urban pollution (Moltchanov et al., 2015). Two of ~~Moltchanov and colleagues'~~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 ~~They found variations in microenvironmental pollutant concentrations, but lacked~~ 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).~~Two of their sites (A and B) had correlations between 0.82 and 0.94 with each other, but correlations between A or B and C were much lower, between 0.04 and 0.72. Mead and coworkers established static and mobile air quality sensor networks in Cambridge, UK measuring CO, NO and NO_2 with electrochemical sensors while performing similar campaigns in Spain and Nigeria (Mead et al., 2013).~~ In 2013, Williams ~~and coworkers~~et al. quantified a tungstic oxide ozone sensor in the lab while addressing some of the main drawbacks associated with m~~M~~etal oxide (MO_x) ozone sensors (i.e. drift/long term stability, material degradation and sensitivity fluctuations) (Williams et al., 2013). The ozone sensors in that~~is~~ 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 ~~this study~~the work presented here, temperature ~~is~~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 ~~and coworkers~~et al. 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 (Lin et al., 2015). 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 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 ~~area~~ frequently experiences ~~very~~ 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~~s~~ is ~~ideal~~ 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 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. This research was conducted in part for the CyberSEES project (NSF Award ID: 1442971).

2 Methods

This field study was conducted within a ~~314~~200 km^2 area of n~~N~~orthwestern Riverside c~~C~~ounty, 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 monitors were deployed within a 8~~10~~ 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 the South Coast Air Quality Management District (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 neighborhoods 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 ~~are a high~~is a large number of vehicles traveling around and through this study area daily; these vehicles likely represent the dominant sources of ~~CO_2,~~ NO_x, and VOCs, precursors to ozone formation.



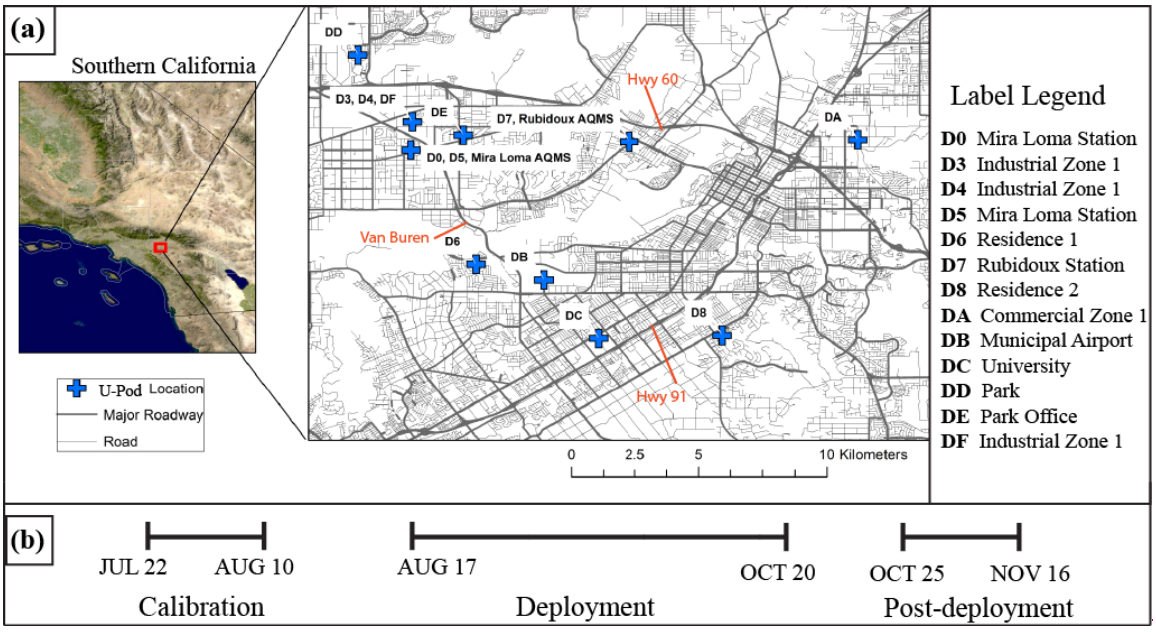


Figure 1. (a) A ~~m~~Map of the deployment area. The crosses indicate U-Pod locations, with the AQMS labelled by name and (b) a timeline of project phases, from calibration ~~,~~ to deployment. Validation overlapped with the deployment time period. ~~and post-deployment~~

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, ~~MO_x,~~ (MiCS 2611, SGX Tech. formerly e2v ~ $11), ~~, CO_2 is measured with a low-cost non-dispersive infrared (NDIR) sensor (S200/300, ELT Corp. ~$50).~~ 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 ~~innate~~ 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. ~~A detailed~~Comprehensive ~~summary~~ reviews of MO_x gas sensors (Korotcenkov et al., 2007) and experimental tests (Masson et al., 2015; Rai et al., 2017) ~~by Korotcenkov and coworkers~~ document~~s~~ potential ~~issues~~ concerns of using sensors in long term ambient monitoring campaigns ~~(Korotcenkov et al., 2007)~~ 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~~s~~ 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~~ing~~ 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.

~~However, due to the low-cost nature of such sensor technology, the use of MO_x gas sensors in ambient monitoring is being widely pursued. NDIR sensor technology uses optical absorption in the IR band to measure CO_2 gas concentration, among other compounds. Most commercialized NDIR sensors commonly contain low-cost mini-filament bulbs (as opposed to LEDs), as do the ELT Sensor Corp. S-300 and S-300A models of the CO_2 sensors used in this study (Bogue, 2013). These sensors operate between -10 and 60 degrees Celsius. They can perform in relative humidity levels between 0 and 95 percent. More information about these specific carbon dioxide sensors is available at http://www.eltsensor.co.kr/2016/products/oem_modules/S-300.html.~~

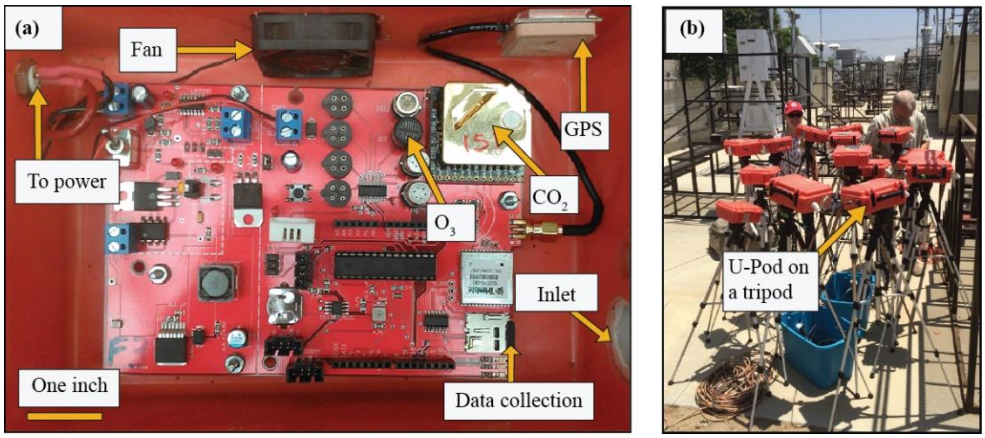


Figure 2. ~~(a)~~ Demonstration of the U-Pod layout (a), including sensor locations and other features. (b) A photo of the field calibration collocation at Rubidoux AQMS.

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 relationships regressions between the reference measurement and pollutant 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. CO_2 was measured using an infrared LI-840a gas analyzer (LI-COR) which was lab-calibrated at the University of Colorado, using three certified gas standards (0, 287 and 1990 ppm) prior to the campaign. 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. For the NDIR carbon dioxide sensors, the signal is an analog output. 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. Through this process, we discovered that an 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 (Eq. 1).

$$S = p_1 + Cp_6(T + p_2) + TAp_3 + Ap_4 + (t - t_o)p_5 \quad (1)$$

Where In Equation 1, S is the sensor signal in R/R_o, where R is the sensor resistance and R_o is a chosen specific normalizing resistance value. C is the pollutant concentration in ppbv or ppm, T is the temperature in Kelvin, A is absolute humidity in mole fraction, $t-t_0$ is the duration since the start of the calibration and the p variables are coefficients determined by the regression minimising least squares. 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. The values of these coefficients and their physical meanings are described in section 3.1.

For CO_2 data, the best model incorporates the same variables as Eq. (1), but in a slightly different combination (Eq. 2).

$$S = p1 + Cp_2 + Tp_3 + Ap_4 + (t - t_o)p_5 \quad (2)$$

2.3 Field Validation of Model Performance

To quantify the performance of the calibration model coefficients, aA 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. to quantify the performance of the generated calibrations in the field. 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 and carbon dioxide sensor measurements to reference grade monitors at one-minute resolution. Two validation approaches were investigated here. 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 that was different 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.outside the calibration microenvironmental space . 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 ~~methods~~ is presented in the results.

3 Results

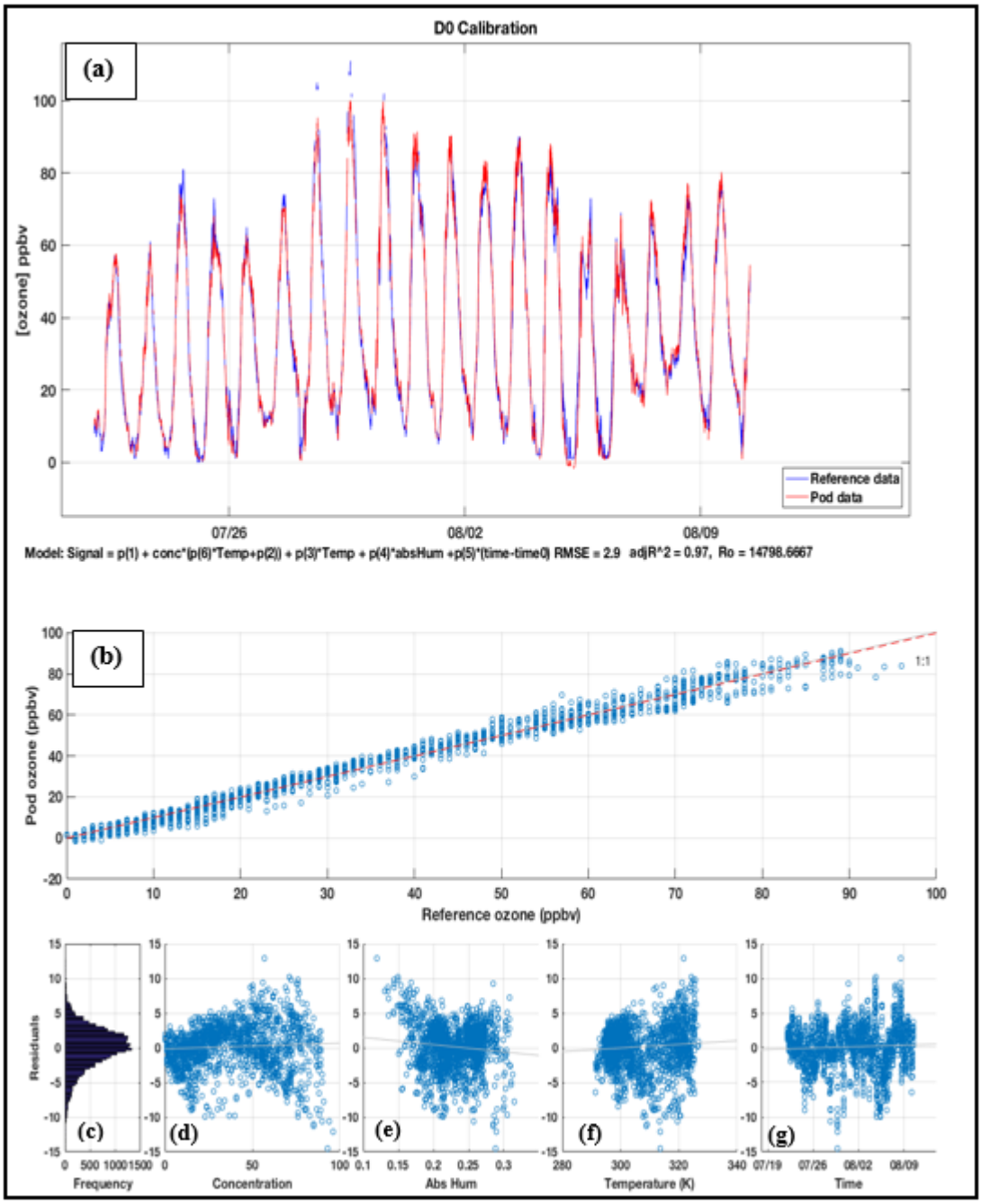
3.1 Field Calibration Results

~~Results of the field calibration process are provided in Table 1 and sample models are shown in Fig. S1 and S2 (for ozone and carbon dioxide, respectively).~~ 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 O_3,~~ For the sake of simplicity, results from ~~the 4T model (Eq. 1),~~ ~~t~~the overall best performing model, see Eq. 1, ~~statistics overall,~~ are shown in Table 1~~were observed for the four term (4T) linear equation~~. 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 ~~4T~~ 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~~4X1~~ illustrates the calibration results for ~~a sample~~U-Pod D0~~ D0~~. ~~Panel (a) shows modeled (red) and reference (blue) ozone concentrations while panel (b) shows the resultant scatter plot.~~ Residuals were calculated as ~~(~~modeled ~~conc.~~ minus~~–~~ reference instrument concentrations~~.~~ ~~)~~ The normally distributed residuals shown in panel c were indicative of an unbiased model. Residuals ~~are~~were ~~depicted~~plotted ~~as a function of the~~versus various model parameters to ~~show~~ assess ~~any trends~~bias in the model performance as a function of ~~(via the trend lines) among~~ the predictors. ~~The normally distributed residuals (c) is indicative of an unbiased model.~~ The slightly negative slope of the trend line in panel ~~(~~e~~)~~ indicate~~d~~s under predicting at increasing absolute humidity whereas positive slopes in panels ~~(~~d and f~~)~~ shows the opposite trend, slight ~~(~~over~~–~~prediction~~ng~~ at higher values of concentration and temperature~~)~~. The R^2 and RMSE values for the calibration of ~~D0~~this sample U-Pod were~~are~~ 0.97 and 2.9 ppbv respectively.



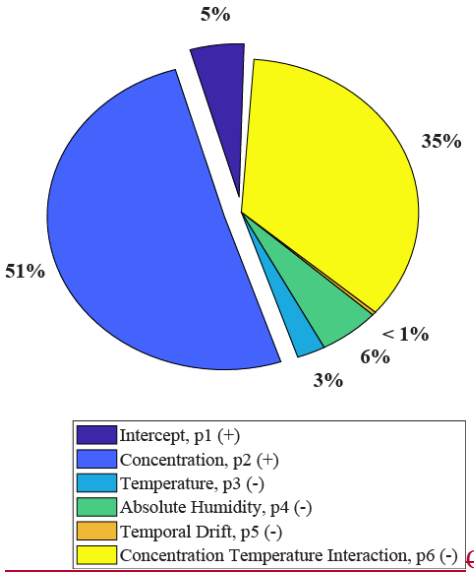
~~There were three other regression equations tested that fit less well; those fit statistics are presented in Table S1.~~

Figure 34~~X1~~. 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 (Barsan and Weimar, 2001) and subsequently Masson et al. (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).

~~Instead of spearheading this discussion here,~~~~Some~~Additional ~~-~~insight into this effort can be gleamed by exploring the results of sensor-specific model parameters~~ are compared to shed light on this topi~~ from ~~this~~the nearly three-week calibration period of this study~~e~~. 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~~5X2~~ ~~below~~ shows the fractional contribution of each model parameter during the calibration period towards estimating the sensor signal (R/R_o)~~the~~ . Concentration (reference, ppbv) and the concentration temperature interaction term combined explain ~~variation~~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 and colleagues 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.



~~CO_2 calibrations were less precise, with R^2 and RMSE ranging from 0.44 – 0.92 and 9.0 – 77 ppmv, respectively (Table S1). Inherent CO_2 sensor auto-calibration likely caused these variations in calibration~~

~~performance because of automated baseline and sensitivity shifts over time, as well as increased noise in this sensor compared to ozone.~~

Figure 4~~5X2~~: 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 -).~~, see Eq. 1,~~

~~variationwas~~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~~was~~. 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).

~~**Table 1: Field calibration results of the Linear 4T model 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,~~	~~0.98,~~	~~0.98,~~	~~0.98,~~	~~0.99,~~	~~0.98,~~	~~0.98,~~	~~0.98,~~	~~0.97,~~	~~0.98,~~	~~0.99,~~	~~0.99,~~	~~0.97,~~	~~0.98,~~
~~RMSE~~	~~3.1~~	~~3.0~~	~~2.6~~	~~2.7~~	~~3.5~~	~~2.8~~	~~3.0~~	~~3.9~~	~~2.8~~	~~2.6~~	~~1.8~~	~~3.4~~	~~3.1~~
~~Sensor 2 R2,~~	~~0.98,~~	~~0.98,~~	~~0.98,~~	~~0.98,~~	~~0.99,~~	~~0.98,~~		~~0.97,~~		~~0.98,~~	~~0.99,~~	~~0.98,~~	~~0.98,~~
~~RMSE~~	~~3.2~~	~~3.0~~	~~2.7~~	~~3.0~~	~~2.4~~	~~3.0~~		~~3.9~~		~~2.7~~	~~1.8~~	~~2.9~~	~~3.0~~

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 rest of the summer~~the deployment. ~~Including those data points in subsequent analysis would be a form of extrapolation that can either be assessed or avoided. As such, the~~ D~~U-Pod d~~eployment data were filtered for ~~environmental variables~~conditions that would require extrapolation, an example of which is shown in Fi~~g.g.~~ 5~~6~~4~~3~~. 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 ~~ozone~~ 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. 5~~6~~4~~3~~a 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. ~~Two of the sensors failed outright, flat lining at some point during the deployment.~~ 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-~~p~~Pod specific.

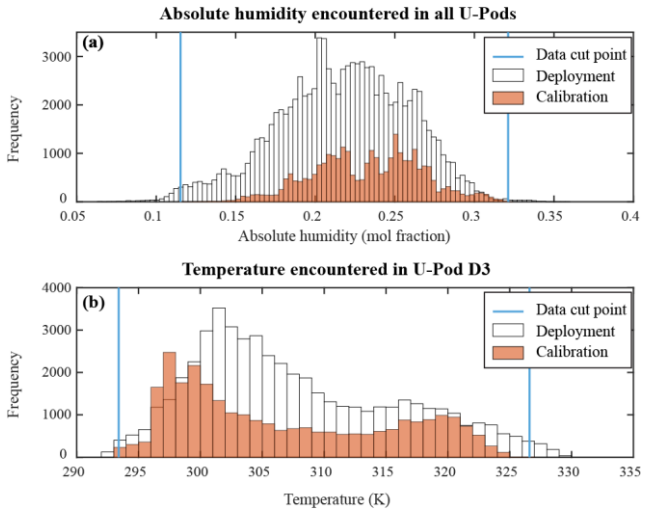


Figure 5~~6~~4~~3~~. Example filtering for a ~~This is an example for one~~ U-Pod (D3) showing lower absolute humidity (a) and higher temperatures (b) occurred during the deployment than during the calibration. The data cut point shows where minimum and maximum values of the variables included in the data were excluded~~are cut.~~ ~~are.~~

In addition, deployment data were filtered for maximum ~~and minimum~~ values of O_3. In some instances, the ozone data spikes to unrealistically high levels. ~~; again, we wanted to minimize model extrapolation.~~ 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. ~~Therefore~~As such, we employed 171 ppbv as~~is~~ a rea~~sonable~~listic maximum level of ozone to~~hat we could~~ expect across the study area. ~~The difference between the maxima of the Rubidoux and the Mira Loma air quality monitoring stations was 2 ppb for the deployment period, or 7 % of the highest maximum value recorded by either station. For O_3, values that resulted in~~ Ceoncentrations that were over ~~7% of the highest maximum value~~this threshold ~~(148 ppb)~~ were removed. No minimum filtering was needed for O_3. ~~For CO_2, there was no maximum or minimum filtering.~~

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 are shown in Table S2. Most U-Pods (except D8 and DB) have two ozone sensors~~., and one CO_2 sensor.~~ 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.

~~Subsequently,~~ 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 ~~to begin with~~ 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. ~~CO_2 sensor, with the goal of disabling the auto-calibration setting. For the ELT-300 model, this auto-calibration happens during "dimmed" light settings for indoor monitoring, creating challenges for ambient monitoring.~~ This ~~electrical~~ 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 ~~from analysis~~ resulted in the loss of three~~two~~ sites from the study. All the remaining sites were left with one U-Pod each.~~The other sites had a least one U-Pod remaining.~~

3.3 Validation of Field Calibration ~~Validation Results~~

Validation of the f~~F~~ield calibration models was achieved by deploying U-Pods next to reference instruments during times when ~~validation applies to U-Pods during the deployment period that continued to measure at reference monitoring stations while~~ 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 ~~All~~ regression model~~s s~~ (~~shown in~~ Table S1) were applied to the filtered ~~validation datasets~~data from D7, D0 and D5. The best performing model was selected based on R^2, RMSE and residual distributions. Ozone concentrations were best model~~l~~ed over the entire validation time period using the ~~linear4T~~ 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. ~~Carbon dioxide concentrations were best modelled using a linear relationship with sensor signal, absolute humidity and temperature, dubbed Linear3 (Eq. 2).~~ 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 r~~R~~esults from the entire validation period ~~both field validation methods~~ are presented in Figs. S2~~3~~, S3~~4~~ and S4~~5~~ for pods D0, D5 and D7, respectively.~~, and summary statistics are shown in Table 2.~~

Table 2. Overall v~~V~~alidation sensitivity results showing mean residuals, median residuals, R^2 and RMSE of sensor measurements against Rubidoux or Mira Loma AQMS O_3 (ppbv) observations ~~for O_3 (ppbv) and CO_2 (ppmv).~~ Two-hundred iterations of 10% ~~of~~ randomly ~~chosen~~selected ~~data~~minute-data ~~was~~were used for validation statistics (~~–~~± 1 SD).~~The residuals are U-Pod data – reference monitor data.~~

U-Pod ID~~/species~~	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

We thought that the The first validation method (U-Pod in the same location as the reference station, D7) would 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 U-Pod D0 show better or similar performance to the Mira Loma station reference data than the two sensors in U-Pod 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-Ppod 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 67X3.

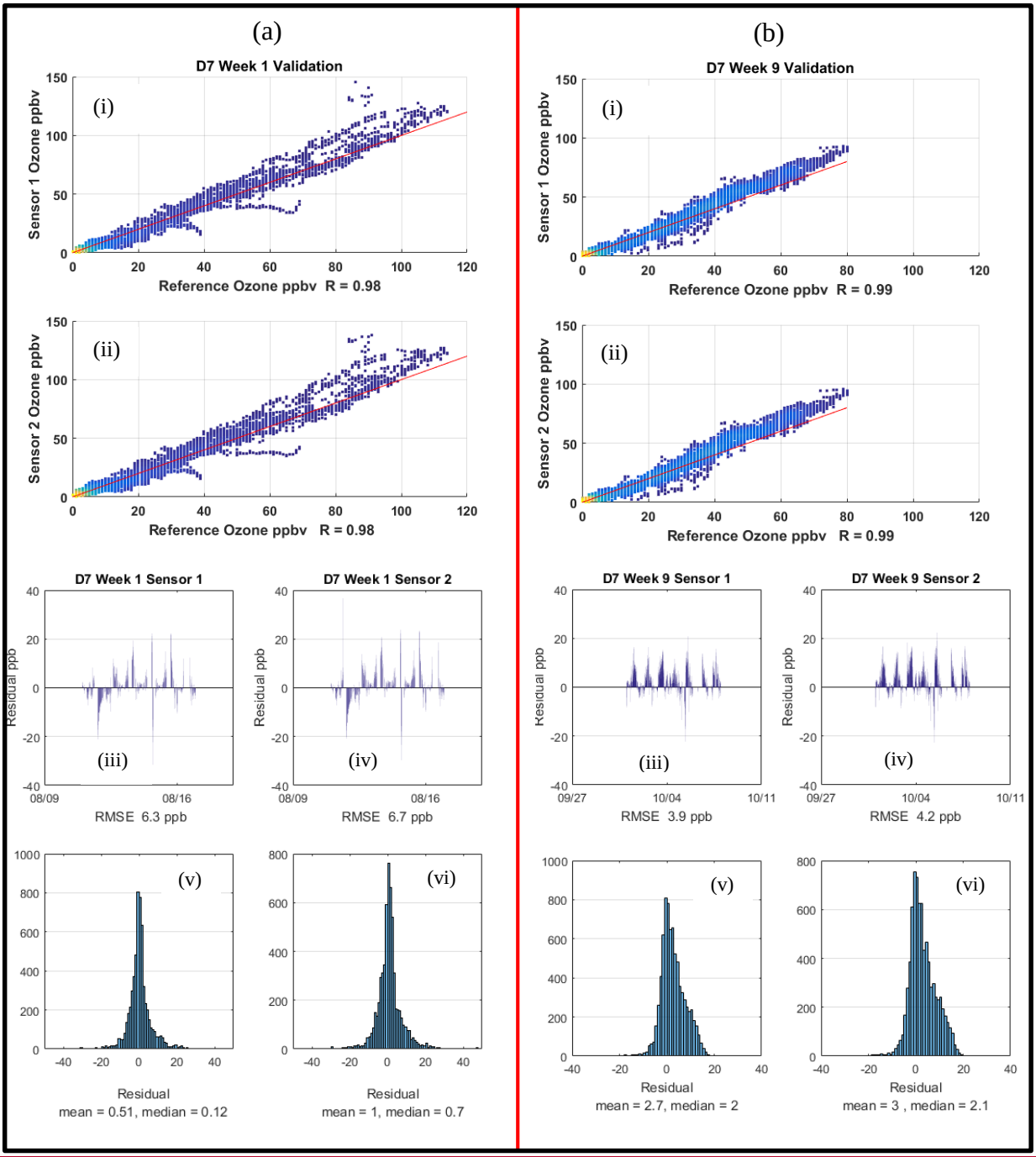


Figure 576: Validation results from the (a) first week and (b) ninth week of the deployment period for D7 ozone sensors separated by the red line. Subpanels (i) show a scatterplot of sensor 1 and reference measurements with warmer shading showing a higher density of points, (ii) show a scatterplot of 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~~7~~5 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.~~ure~~ 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 ~~of ozone~~ between 90-1~~—1~~00 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. ~~This corresponds to a concentration of about 5 ppb.~~ Values that are ~~less than~~within ~~around~~ 5-~~%ppb~~ of the standard~~different than expected,~~ would not be flagged. This serves as a reference point for the quality of the reference ozone measurements. During ~~the~~ validation, ~~our~~ O_3 sensors had ~~a~~ measurement error (RMSE), median residual and mean residual ranges of ~~for minute median data, which ranged from~~ 4.3 – 7.3~~4.45 to 7.16 ppb. For each sensor, the median of the residuals ranged from 0.86 to 7.15 ppb, while mean of the residuals ranged from 0.7 to 6.4 ppb~~, 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, ~~U-Pod~~ 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 b~~B~~ias for the other U-Pods was relatively small and on the order of 1-~~—~~2 ppbv.

~~For CO_2, the RMSE was 15.0 ppmv. The sensor was under predicting at higher concentrations and higher absolute humidity, and slightly over predicting at elevated temperatures and as the deployment progressed (Fig. S6).~~

~~In order to gain a better understanding of the dependency of model performance on the selection of the validation data, we repeated this assessment with 200 iterations of 10% randomly sampled minute level deployment data. The values of the statistics could be affected by which time period the validation data is selected from. For example, choosing 10% of consecutive data from the first few days could give misleading statistics. The resulting distributions for the performance metrics are shown in Table 2. The +/- indicates the range of values observed during all the iterations. Tight distributions show little dependence on the data selected.~~

~~Table 2. Validation sensitivity results showing mean residuals, median residuals, R^2 and RMSE of sensor measurements against Rubidoux or Mira Loma AQMS observations for O_3 (ppbv) and CO_2 (ppmv). Two hundred iterations of 10% of randomly chosen data was used for validation statistics. The residuals are U-Pod data – reference monitor data.~~

U-Pod ID/species	mean residual	median residual	mean R^2	mean RMSE
D7 O_3 Sensor 1	2.4 +/- 0.1	1.2 +/- 0.1	0.965 +/- 0.001	5.6 +/- 0.1
D7 O_3 Sensor 2	2.8 +/- 0.1	1.5 +/- 0.1	0.963 +/- 0.001	5.9 +/- 0.1
D0 O_3 Sensor 1	0.7 +/- 0.1	0.8 +/- 0.1	0.974 +/- 0.001	4.4 +/- 0.1
D0 O_3 Sensor 2	1.1 +/- 0.1	1.0 +/- 0.1	0.971 +/- 0.001	4.9 +/- 0.1
D5 O_3 Sensor 1	5.5 +/- 0.1	5.1 +/- 0.1	0.971 +/- 0.001	5.0 +/- 0.1
D5 O_3 Sensor 2	6.4 +/- 0.1	3.9 +/- 0.1	0.953 +/- 0.001	7.2 +/- 0.1
D7 CO_2	3.0 +/- 0.3	3.0 +/- 0.3	0.789 +/- 0.007	15 +/- 0.2

3.4 Deployment Data

As mentioned above, U-Pods were deployed, spread out across 200 km^2 2 area in Riverside, CA; as such, This section concerns the analysis of all the U-Pods during the deployment period, including the ones used for validation. Thethe 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 101 week deployment were used in the following analysis. The calibration model coefficients obtained during collocationthe calibration time period (collocation with the reference monitor) was were applied to all data during both the calibration and deployment time periods. Applying the linear model to the U-Pod 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 and deployment data are useful to assess as the "variability" observed in the collocation data approximates that measurement uncertainty, and as such comparing that "variability" to the variability observed when the U-Pods are deployed vs. when they are collocated. This allows us to investigate our ability to observe actual spatial and temporal differences. In all following figures, hours of the day are given in local time.

3.4.1 Ozone

The U-Pods sampled for approximately 2900 hours total, 58% of which consisted of the deployment period data. The medians of ozone value distributions by U-Pod during the calibration range from 29 to 30 ppbv. During calibration, t. The 5th% percentile and 95th percentiles ranged from values range from 2 5 ppbppbv and and the 95th percentiles were 70 83 ppbv, respectively.

During deployment, the median ozone ~~distribution~~ values were between 14 and 31 ~~ppb~~ppbv while ~~.~~ ~~t~~The 5th ~~percentile~~ and 95th percentile ranges were ~~range for deployment was~~ 0~~.~~ ~~–~~6 ~~ppb~~ppbv and ~~while the 95th percentile was~~ 67~~.~~ ~~–~~99 ppbv~~b~~, 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 and co-workers 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 8~~4~~ shows the diurnal cycle for ozone based on concentrations collected during this study.

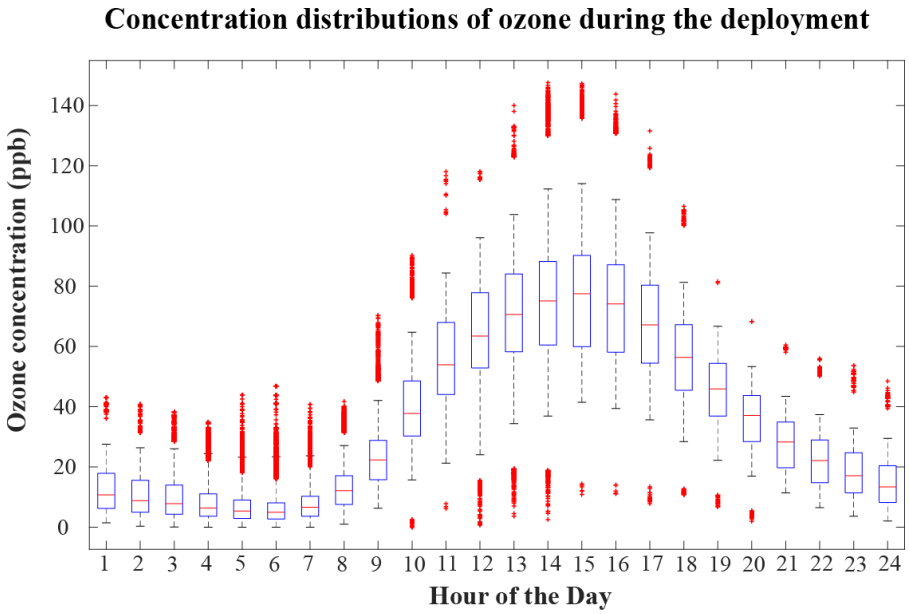


Figure ~~4~~7~~8~~. The diurnal cycle of ozone during the deployment. Distributions are concentrations from all U-Pods during each hour. Whiskers indicate the 5th and 95th percentile, with + marks falling outside of this range. The box boundaries span the 25th to 75th percentiles.

Figure 7~~8 4 gives some~~offers context of what the temporal variability in ozone concentrations in this study looks like. There are ~~some~~ trends in ozone concentrations across ~~the~~ Southern California that ~~we would expect to see~~would be expected. ~~Ozone generally is inhibited in the early morning, followed by midday growth and a peak in concentration. For the remaining hours, ozone concentrations generally decrease, although emissions from the previous day may have effects on concentrations during the next (Gao and Niemeier, 2007)~~Ozone~~. For this study, 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, w~~W~~e examined the R^2 values for all possible U-Pod pairs for each hour of the day ~~to understand spatial variability~~. 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~~9~~5 shows ~~this~~ correlation information between U-Pods for each hour of the day for ozone. For this plot, all ~~U-Pod~~ 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.

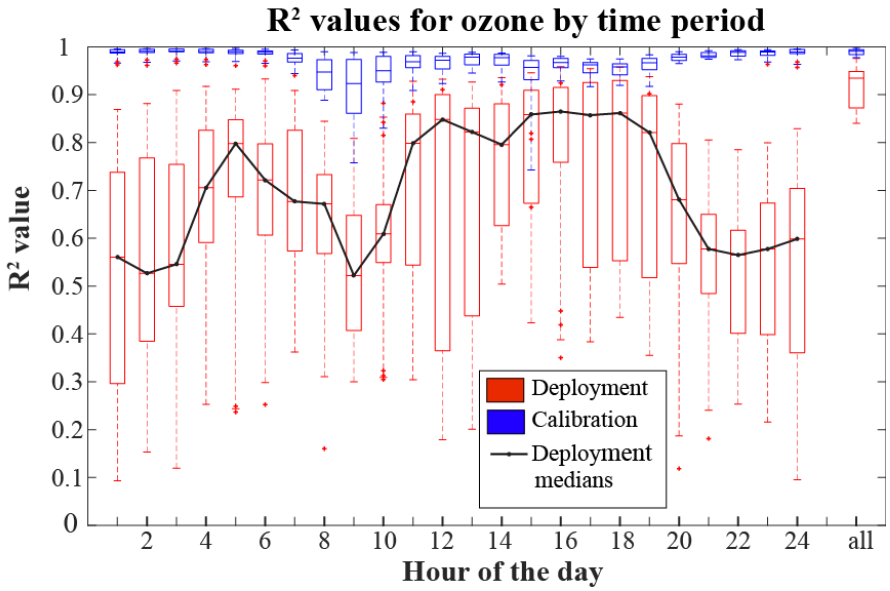


Figure 8~~9~~5. Each boxplot is a collection of the R^2 values between every pair of U-Pods for each hour of the day. There are 21 points in each boxplot. Medians of distributions are marked by horizontal lines. Whiskers indicate the 5th and 95th percentile, with + marks falling outside of this range. The box boundaries span the 25th to 75th percentiles. The "all" category includes all hours of the day.~~U-Pod~~

~~The~~ U-Pod ozone measurements ~~s~~ 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~~9~~5 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~~10~~6 shows distributions of median absolute differences. All the minute median ~~U-Pod~~ data ~~was~~ were time-matched and binned by ~~the~~ 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 ~~were~~ was calculated. The distributions in Fig. 9~~10~~6 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.

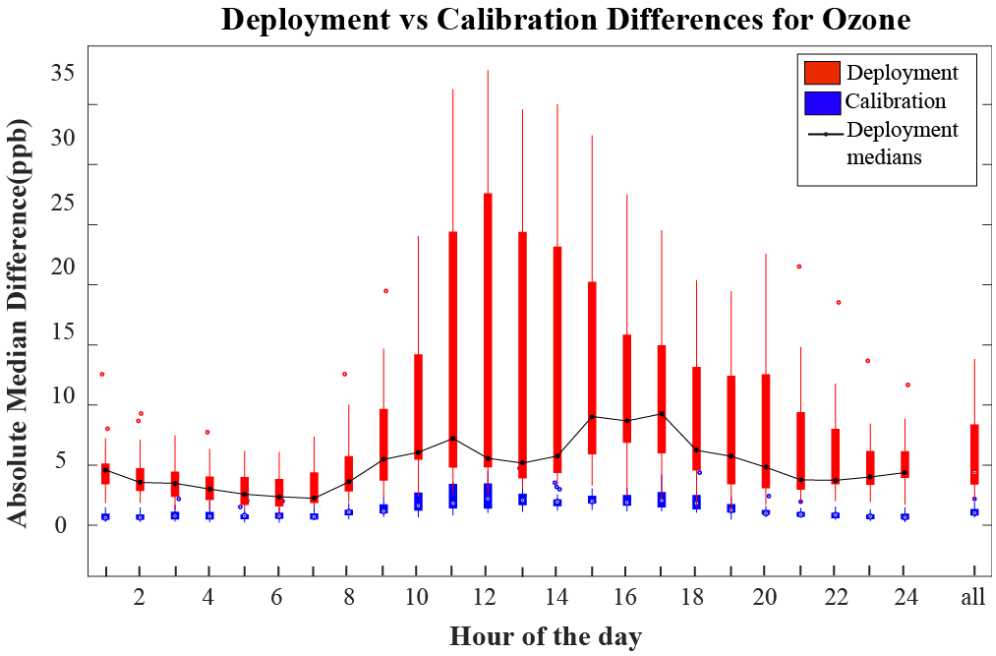


Figure 9106. Distributions of medians of absolute differences between all pairs of pods for each hour of the day. Whiskers show 95% intervals. The black line connects the medians of the deployment. The "all" category includes all hours of the day.

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. 8984 and 91095 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 Hhowever, this inverse relationship between large R^2 and smaller differences does not appear. tThe 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. 9106 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. ure 74 shows high concentrations during the day, Fig.ure S65 A1 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-Ppods 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, indicated 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 9106 shows that the range of spatial absolute differences in O_3 is smallest at night. However, Fig. 895 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 it this assumption could be valid for may apply to other

areas (Moltchanov et al., 2015).~~s.~~ 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 ~~were~~ 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, ~~U-Pod~~ D7 was used for normalization. U-Pod D7 was never moved from Rubidoux station throughout the project and ~~and has a validation set~~as such was employed in the ~~was~~ validation effort mentioned previously~~ed~~. This~~A~~ 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 d~~D~~ifferences between calibration period trends and deployment trends ~~were analyzed~~in Figure 10~~1~~ ~~(Fig. 11~~7~~)~~ as well as hourly ~~trends by pod~~patterns in Figure 11~~2~~. ~~(Fig.s. 7 and 12~~8~~).~~

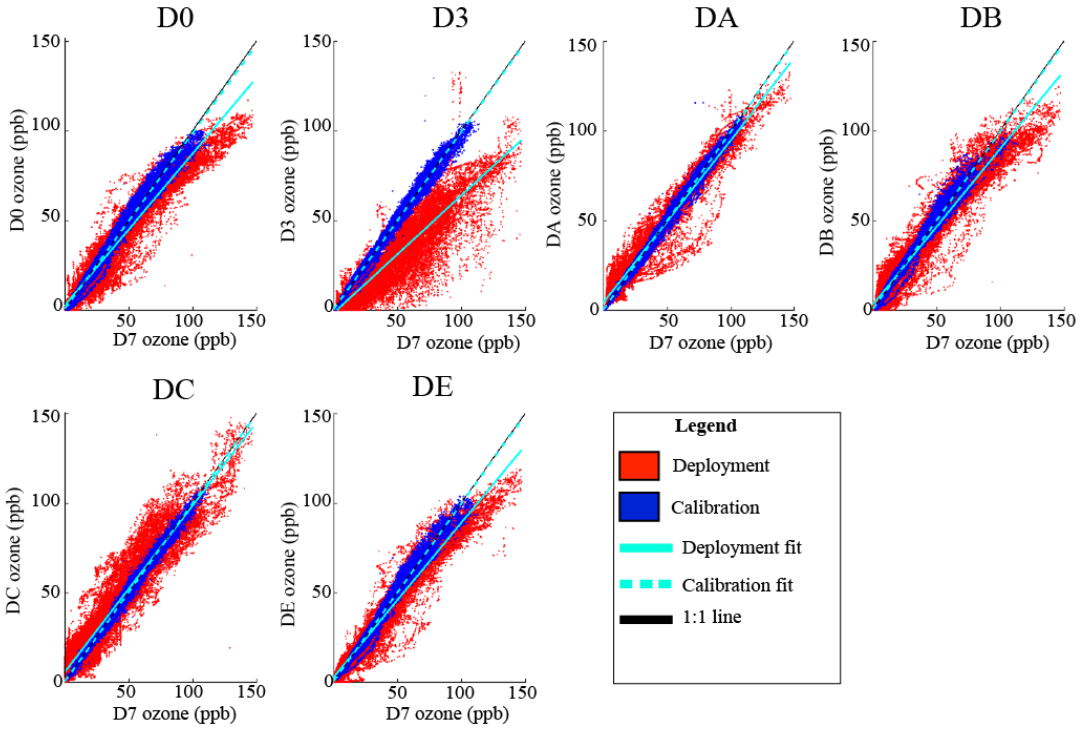


Figure 10~~1~~7. U-Pod D7 ozone concentrations are plotted on the x-axis and other U-Pod ozone concentrations recorded at the same times are on the y-axis. The sets are color coded according to time period their data were taken, and each color is fit with a linear line.

In Fig. ~~7~~10~~1~~, 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. ~~U-Pods D0, D3 and DE are quite close to each other (< 2km) and close to Van Buren Blvd. The deployment O_3 concentrations for U-Pod D3 features a trend that differs significantly from its calibration set.~~ 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, ~~U-~~

~~Pod~~ D7 at the Rubidoux site typically measured higher O_3 ~~O3~~ 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~~east of the project area, and all were less than 1.2 km from the road. Therefore, one might expect ~~the~~ data from these U-Pods to be very similar. Indeed, D0 and DE have similar data cloud shapes in Fig. 10~~1~~7. However, ~~the~~ data from ~~the U-Pod at~~ 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, t~~T~~he lower ozone values of D3 compared to D7~~,~~ also appears more pronounced on weekdays (Fig. S7~~6~~A~~2~~) 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 ~~a~~ road~~s;~~ DC is closest (0.6 km) to ~~the major roadway,~~ 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~~2~~8 and ~~9~~123 demonstrate this ~~idea~~concept.

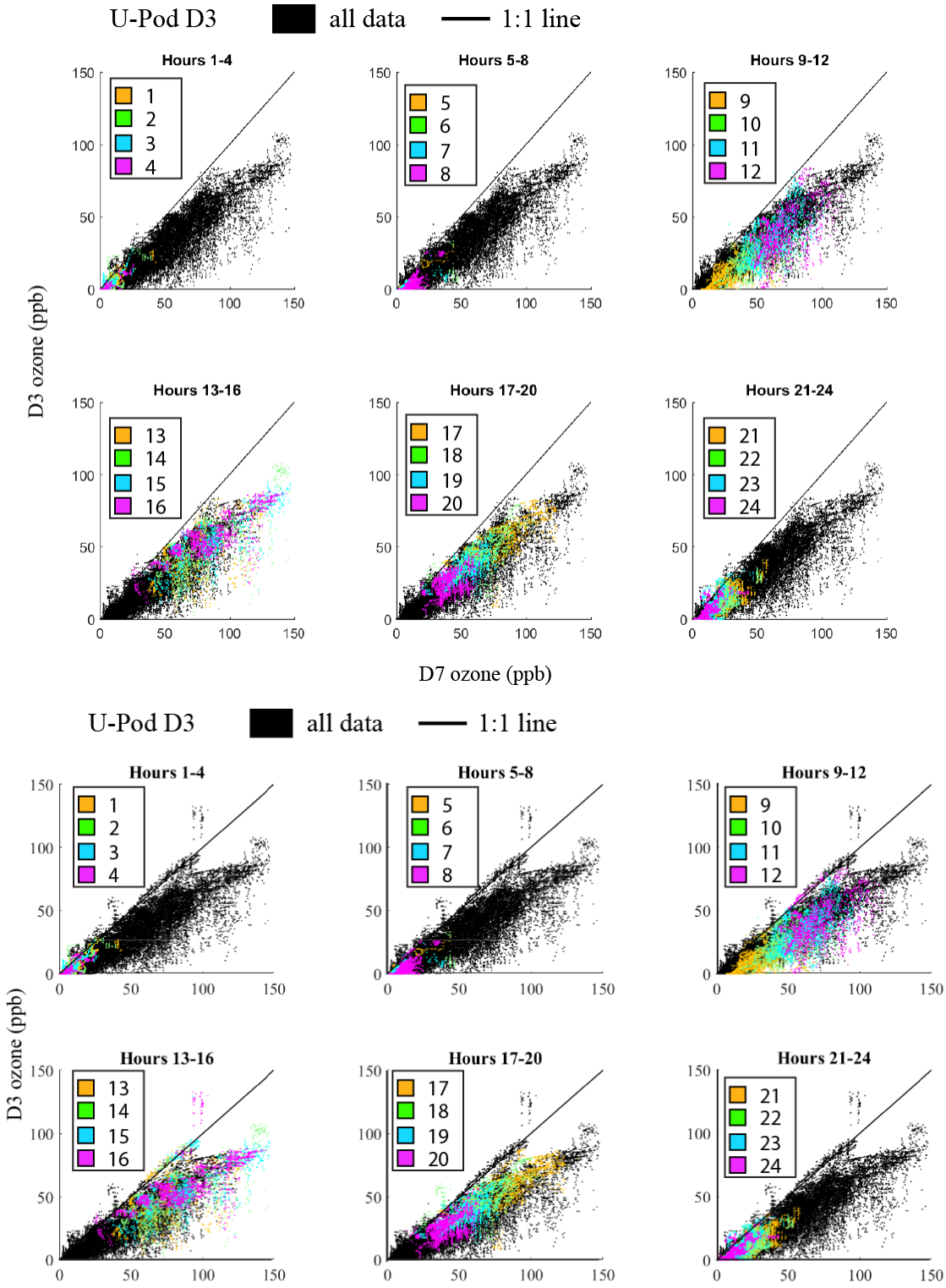
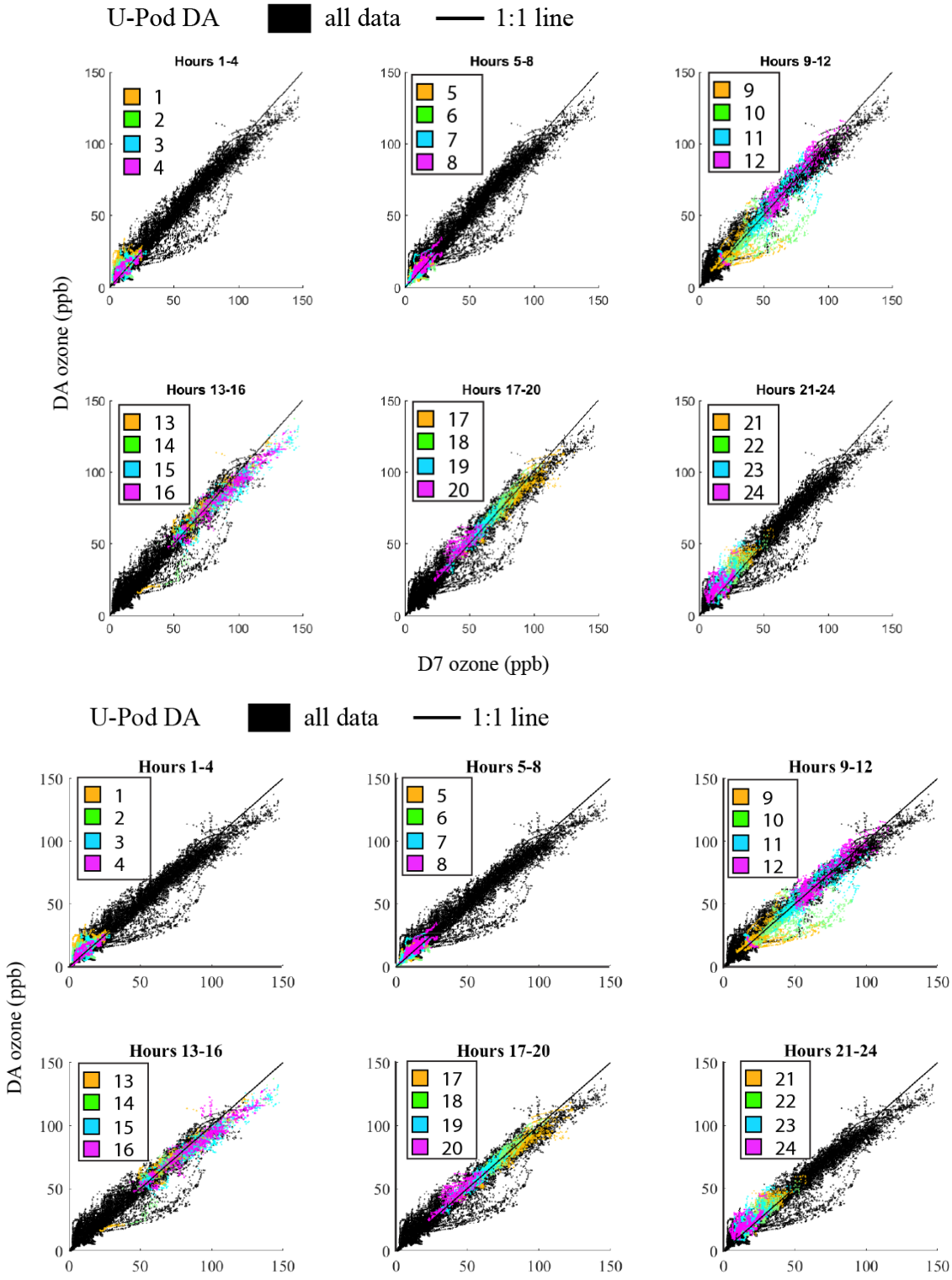


Figure 11~~2~~8. ~~This plot is for~~ Data from~~U Pod~~ D3, at Industrial Zone 1, plotted against D7 (at Rubidoux). In e~~E~~ach scatterplot, colored data in the legend represents ~~is~~ four hours of the day, ~~with~~ and the black data represent~~sing~~ the complete deployment dataset (all hours). The black line is a 1:1 line, not a line of best fit.

~~and data points recorded within each hour bin are marked by the colors and times in the legend.~~

Figure 10~~17~~ and 11~~29~~ 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. ~~Figure 8 shows the comparison of D3 with D7 by time of day.~~ U-Pod D3 was sited at a company in an industrial area where there are potentially more VOCs in the air. ~~It is also~~This site was half a kilometer from the Van Buren roadway and as such ~~, so~~ there is also the potential for elevated levels of NO_x. ~~Beginning~~ 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.



U-Pod DA
all data
1:1 line
Hours 1-4
Hours 5-8
Hours 9-12
Hours 13-16
Hours 17-20
Hours 21-24
DA ozone (ppb)
D7 ozone (ppb)

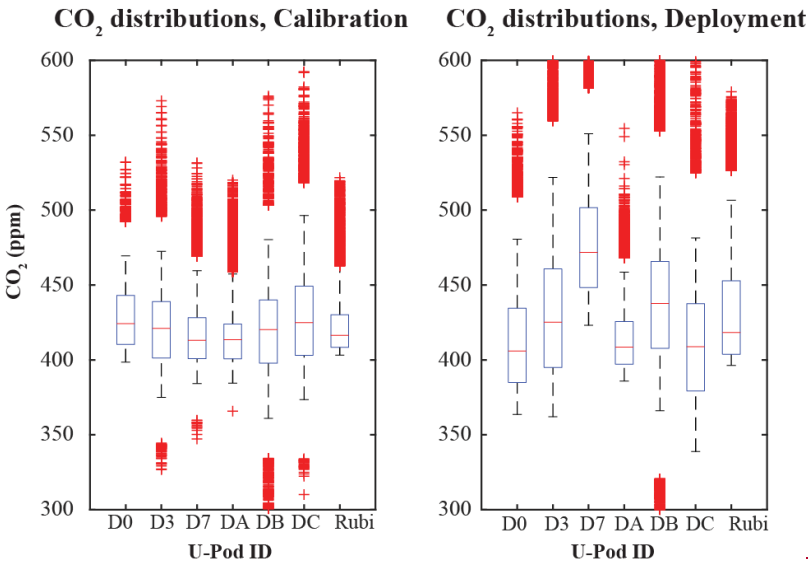
Figure ~~9~~132. ~~This plot is for U-Pod~~Data from DA, located at Commercial Zone 1, plotted against D7 (Rubidoux). Each scatterplot is four hours of the day, with the black data representing the complete deployment dataset (all hours) and data points recorded within each hour bin are marked by the colors and times in the legend. The black line is a 1:1 line, not a line of best fit.

Figure 1239 shows the relationship between U-Pod DA and U-Pod D7 at varying hours during the day.; there were some obvious differences between U-Pods D3 and DA 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. For U-Pod DA, Tthese 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. relative to D7 remain much more highly clustered than in the same hours for U-Pod D3. There was far less spread, which indicates that ozone from U-Pods D7 and DA were more similar than D7 and D3. Also of interest is the strange claw shape on the underside of the black data cloud. The analysis in Fig. 1239 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. So, what is happening at the DA site between 9:00 and 11:00 that was not happening between 7:00 and 9:00 or later in the day? These gases could be The demonstration of variability within this two hour period may be a result of 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. A2S76). 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.

3.4.2 Carbon Dioxide

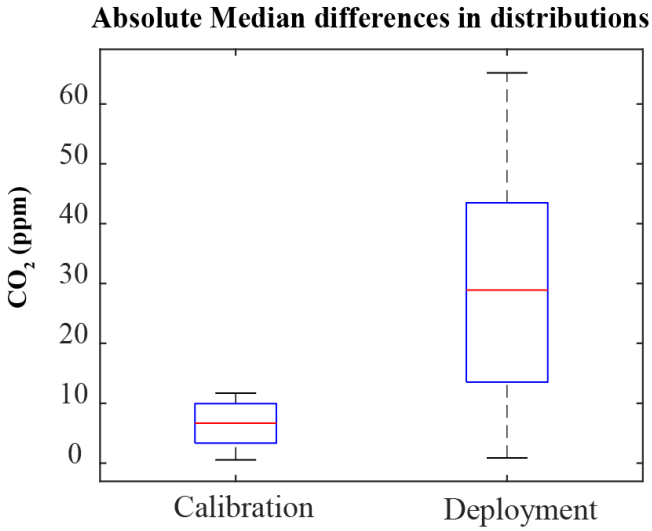
Carbon Dioxide was continuously measured at the same temporal resolution as ozone in the U-Pods over the same time period. CO_2 data from U-Pods D0, D3, D7, DA, DB, and DC were used for this section of the analysis, based on completeness, reasonable baselines, and no obvious trends through time which may signal sensor drift.

To explore the ability to detect spatial differences in carbon dioxide, we compared the distributions of measurements between the calibration time period and that of the deployment, as shown in Fig. 10.



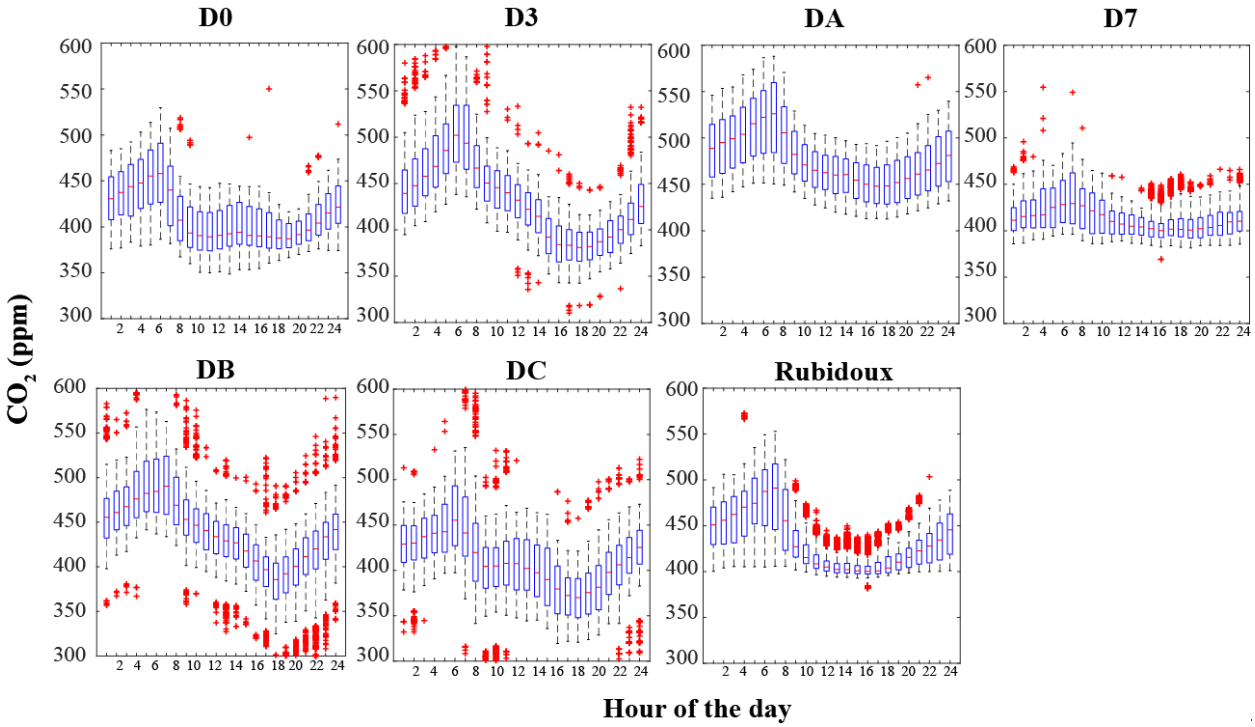
~~**Figure 10. Distributions of carbon dioxide measurements from the (a) calibration and (b) deployment time periods. Whiskers indicate the 5th and 95th percentile, with + marks falling outside of this range. The box boundaries span the 25th to 75th percentiles. "Rubi" is shorthand for carbon dioxide concentrations measured at Rubidoux AQMS with a LI-840a gas analyzer, which is not managed by SCAQMD.**~~

~~The medians of each pod carbon dioxide distributions vary from 413 – 425 ppm during the calibration and 406 – 472 ppm during the deployment. Several differences in the distributions are visible between the two time periods. U-Pod D7 observed higher carbon dioxide measurements during the spatial deployment as compared to the collocated calibration whereas some U-Pods, like DA, have more similar measurements across the two. To more clearly see the variability of measurements between the two time periods, absolute differences between U-Pods were examined. Figure 11 shows the median absolute differences for every possible pair of U-Pods in two groups, calibration and deployment. The differences between U-Pods when they are collocated were small compared to difference between U-Pods when they are dispersed. Some of these differences are greater than the uncertainty of 15 ppm, demonstrating spatial variability.~~



~~**Figure 11. absolute differences of distribution medians over the two time periods: calibration and deployment.**~~

~~We further examined the spatial CO_2 differences by comparing hourly CO_2 concentrations at each site, see Fig. 12. Carbon dioxide values are affected by anthropogenic, biogenic, and atmospheric phenomena. The similar peaks and troughs in the trends likely corresponded to changes in height of the boundary layer. A shorter boundary layer during nighttime would concentrate pollutants, while a larger boundary layer would dilute them. Most of the minimum median CO_2 concentrations occur around 16:00, which is consistent with another LA carbon dioxide study (Newman et al., 2013).~~



~~**Figure 12. Hourly distributions of carbon dioxide over the deployment period, including the reference data.**~~

~~The distributions shown in Figure 12 were similar in several ways. Between 6:00 and 8:00, carbon dioxide concentrations were at their peak, and the lowest values occur between 16:00 and 19:00. However, some U-Pods had different distributions in terms of range of data and apparently shifted baselines. The carbon dioxide at Rubidoux station almost never recorded data below 400 ppm, while all the U-Pods except DA do. Also, there appear to be two types of shapes of hourly data. D0, DA, D7, and Rubidoux belong to one group, in which the minimum and maximum medians are closer together. D3, DB, and DC all have minimum values that are much lower than the early morning peaks. The pods that have more similar trends are not closer together than the others. Differences in the trends between hours might be related to traffic counts, or amount of vegetation. Further investigation is warranted.~~

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~~300~~ km^2. Field validation of sensor O_3 ~~and CO_2~~ measurements to minute resolution reference observations resulted in R^2 and RMSE of 0.95~~—~~0.97 and 4.4~~—~~7.2 ppbv ~~for ozone and 0.79 and 15 ppmv for carbon dioxide, respectively~~. 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 ~~ppb~~ ppbv for ozone. The measurements from the MiCS 2611 ~~O3~~ ozone sensor should not be thought of as a way to replace regulatory ~~air quality monitoring stations~~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. ~~This work takes more time to perform, and requires more people than more expensive, but user-friendly regulatory procedures and equipment.~~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.

~~For ozone analysis, the~~The ~~data show the~~ highest amount of variability between U-Pods based on the R^2 ~~squared~~ 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 ~~p~~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~~-~~ ~~1~~6:00. The uncertainty of these measurements, as determined by the validation results of D0 and D7 is 4.4 – 5.9 ppbv.

~~For CO_2 trends, it appears that this S200/300 CO_2 sensor from ELT Corp. auto-calibrating feature caused issues for characterizing the sensor signal over time leading to issues applying calibration models to raw data. U-Pod CO_2 data from the calibration period do not have as high of correlations with each other as those from the ozone data. However, it is still possible to see differences greater than their uncertainty between U-Pods during the deployment, indicating measureable heterogeneity across the study area.~~

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,~~Take this study for example,~~ 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. I~~i~~t is important to collocate the sensors ~~frequently~~ as frequently as possible while balancing other resources. ~~in order to adjust the linear regression coefficients, and therefore provide a better estimate of values throughout the whole deployment. Another area to examine for S~~sensor quantification ~~could be~~using different mathematical approaches to ~~the~~ 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 ~~a~~ non-linear models.

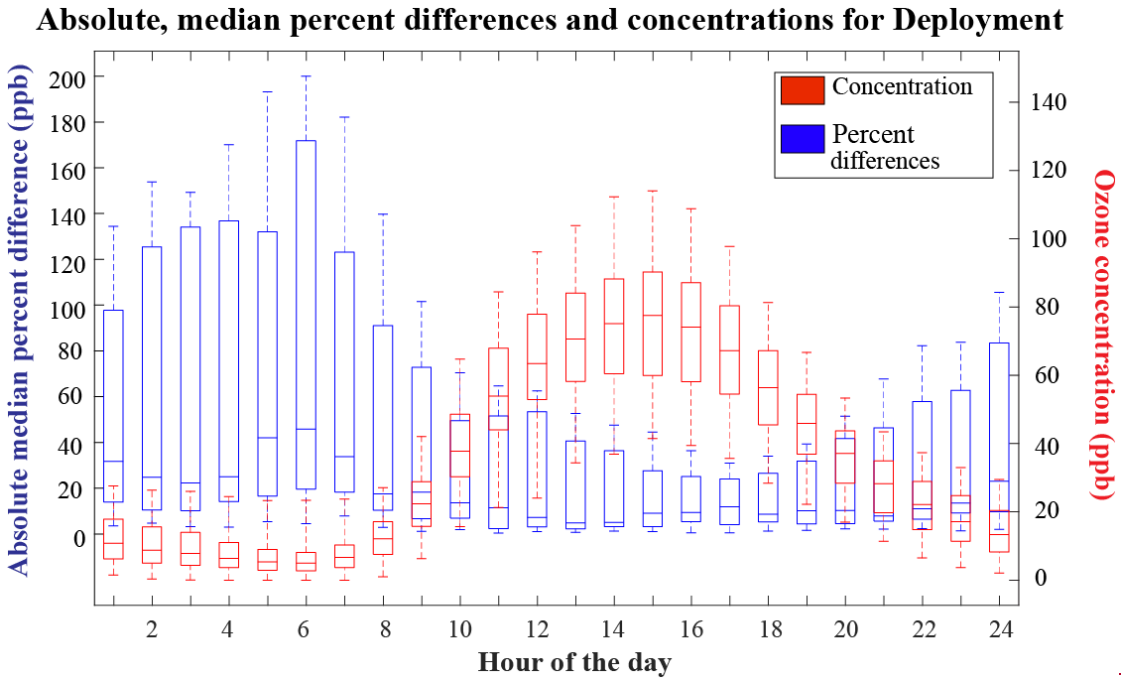
Additionally, including measurements of other compounds in the study could help to explain causes for spatial and temporal variability in both ozone ~~and carbon dioxide~~. For example, including information on nitrogen oxides could help inform the effects on traffic on these compounds, while land use 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.

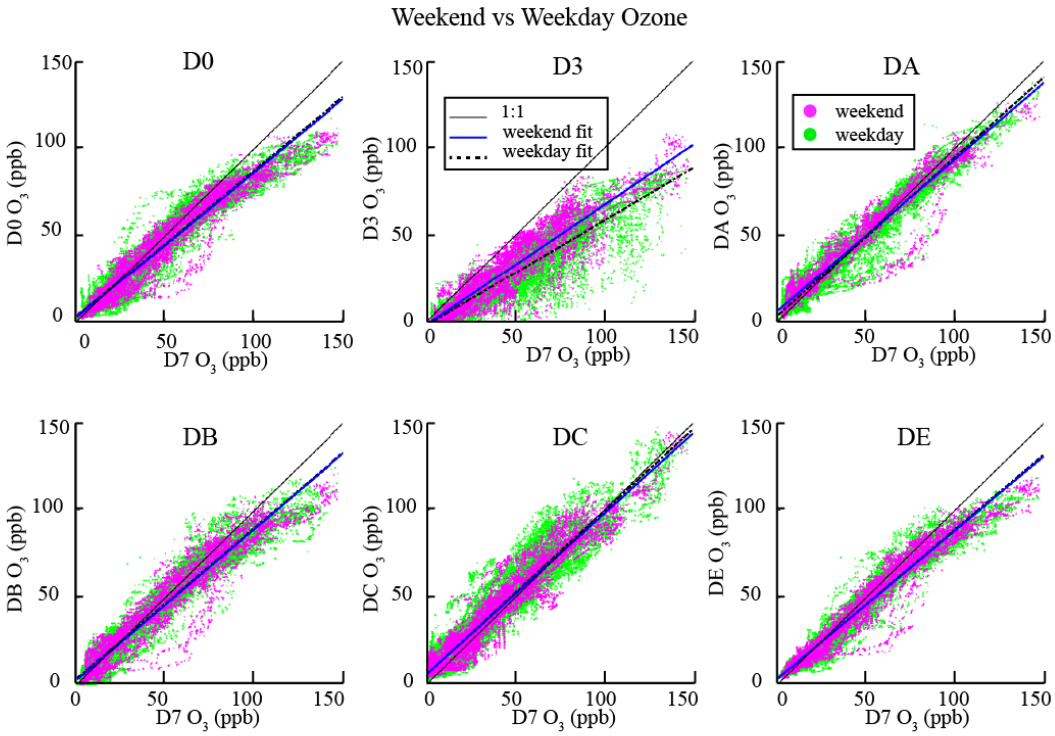
Code and Data availability

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. ~~It would require a lot of extra work from co-authors to guide a third party through their use.~~ 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~~ne~~ usual procedures of quality assurance and quality control before they were provided to us. ~~Also, SCAQMD did not operate or maintain the LI-840a gas analyzer that collected the CO_2 reference data.~~

~~Appendix A~~



~~Figure A1. The right-hand axis shows the distribution of ozone concentrations from all U-Pods for each hour of the day over the deployment. The left-hand axis is the relative percent differences in concentration between all possible pod pairs. Percent difference is used here as the difference in concentration between two U-Pod pairs, normalized by their average. Whiskers indicate the 5th and 95th percentile, values outside of this range are not shown. The box boundaries span the 25th to 75th percentiles.~~



~~**Figure A2. During the deployment period, the magenta data represents data points recorded on the weekend, while green data was recorded during the week. Each subplot is a different U-Pod compared to U-Pod D7 ozone.**~~

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. 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 ~~tech~~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).

Bogue, R.: Recent developments in MEMS sensors: a review of applications, markets and technologies, Sensor Review, 33(4), 300–304, doi:10.1108/SR-05-2013-678, 2013.

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.

~~Gao, H. O. and Niemeier, D. A.: The impact of rush hour traffic and mix on the ozone weekend effect in southern California, Transportation Research Part D: Transport and Environment, 12(2), 83–98, doi:10.1016/j.trd.2006.12.001, 2007.~~

Hutyra, L. R., Duren, R., Gurney, K. R., Grimm, N., Kort, E. A., Larson, E. and Shrestha, G.: Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective, Earth's Future, 2(10), 2014EF000255, doi:10.1002/2014EF000255, 2014.

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.

Kort, E. A., Angevine, W. M., Duren, R. and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO2 emissions: Minimum site location requirements for the Los Angeles megacity, J. Geophys. Res. Atmos., 118(3), 1577–1584, doi:10.1002/jgrd.50135, 2013.

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.

~~Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J. J., McLeod, M. W., Hodgson, T. F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J. R. and Jones, R. L.: The use of electrochemical sensors for~~

~~monitoring urban air quality in low-cost, high-density networks, Atmospheric Environment, 70, 186–203, doi:10.1016/j.atmosenv.2012.11.060, 2013.~~

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.

Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L., Lefer, B., Alvarez, S., Rappenglueck, B., Kort, E. A., Andrews, A. E., Peischl, J., Gurney, K. R., Miller, C. E. and Yung, Y. L.: Diurnal tracking of anthropogenic CO2 emissions in the Los Angeles basin megacity during spring 2010, Atmospheric Chemistry and Physics, 13(8), 4359–4372, doi:10.5194/acp-13-4359-2013, 2013.

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.

Verhulst, K. R., Karion, A., Kim, J., Salameh, P. K., Keeling, R. F., Newman, S., Miller, J., Sloop, C., Pongetti, T., Rao, P., Wong, C., Hopkins, F. M., Yadav, V., Weiss, R. F., Duren, R. and Miller, C. E.: Carbon Dioxide and Methane Measurements from the Los Angeles Megacity Carbon Project: 1. Calibration, Urban Enhancements, and Uncertainty Estimates, Atmos. Chem. Phys. Discuss., 2016, 1–61, doi:10.5194/acp-2016-850, 2016.

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

~~**Table 2. Validation sensitivity results showing mean residuals, median residuals, R^2 and RMSE of sensor measurements against Rubidoux or Mira Loma AQMS observations for O_3 (ppbv) and CO_2 (ppmv). Two-hundred iterations of 10% of randomly chosen data was used for validation statistics. The residuals are U-Pod data – reference monitor data with the same time stamps.**~~

~~U-Pod ID/species~~	~~mean residual~~	~~median residual~~	~~mean R^2~~	~~mean RMSE~~
~~D7 O_3 Sensor 1~~	~~2.4 +/- 0.1~~	~~1.2 +/- 0.1~~	~~0.965 +/- 0.001~~	~~5.6 +/- 0.1~~
~~D7 O_3 Sensor 2~~	~~2.8 +/- 0.1~~	~~1.5 +/- 0.1~~	~~0.963 +/- 0.001~~	~~5.9 +/- 0.1~~
~~D0 O_3 Sensor 1~~	~~0.7 +/- 0.1~~	~~0.8 +/- 0.1~~	~~0.974 +/- 0.001~~	~~4.4 +/- 0.1~~
~~D0 O_3 Sensor 2~~	~~1.1 +/- 0.1~~	~~1.0 +/- 0.1~~	~~0.971 +/- 0.001~~	~~4.9 +/- 0.1~~
~~D5 O_3 Sensor 1~~	~~5.5 +/- 0.1~~	~~5.1 +/- 0.1~~	~~0.971 +/- 0.001~~	~~5.0 +/- 0.1~~
~~D5 O_3 Sensor 2~~	~~6.4 +/- 0.1~~	~~3.9 +/- 0.1~~	~~0.953 +/- 0.001~~	~~7.2 +/- 0.1~~
~~D7 CO_2~~	~~3.0 +/- 0.3~~	~~3.0 +/- 0.3~~	~~0.789 +/- 0.007~~	~~15 +/- 0.2~~