Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-183-RC1, 2017

Atmospheric Measurement Techniques Discussions

Interactive comment on "Intra-urban spatial variability of surface ozone and carbon dioxide in Riverside, CA: viability and validation of low-cost sensors" *by* Kira Sadighi et al.

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-

Printer-friendly version

Discussion paper

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.

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.

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"?

In the references for other low-cost sensor experiments, there have been studies using low-cost CO2 sensors: Shusterman et al. (ACP: https://www.atmos-chem-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.

P3 L31: should be "high VOC concentrations". (and presumably NOx?).

P4 L13 "should be there is a large number of vehicles"...

P5 L8: Some indication of why medians were used rather than means? (to reduce influence of extreme outliers?).

P6 L7: semicolon I think should be a colon.

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.).

P6 L18 sentence structure awkward

P6 Eq 2 p1 should have the 1 subscripted.

P6 bottom, P7 L1: Martin et al. did this for CO2 sensors, but only for a much shorter period of time (2 weeks?).

P6 L30 perhaps to add clarification here, note that the comparisons were made be-

tween 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.

P7 L5: microenvironmental space?

P7 L12: this is the first reference of the 4T equation - this is equation 1?

Table 1: I wonder how different the coefficients Px are between different individual sensors? Can the authors give an idea of this?

P7 last line: "during the deployment that were outside the range of those experienced during the calibration time period" might be clearer here.

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?

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?

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?

P9 L24: What does a precision check entail exactly? This should be stated in the text. P9 L25: wording: should say "5% from expected values (corresponding to a concentration of about 5 ppb), subsequent data would be flagged ...". P9 L26 awkward again: "Values within 5 ppb of the expected value would not be flagged".

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?

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. 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?

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).

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.

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.

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?

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?

P12 L13 and later in the text, when examining pair-wise Rˆ2 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?

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?

Figure 5 caption ends with "U-Pod"?

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 ...".

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

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

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 R^2 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.

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?

P15 L16: "as well as hourly trends by pod (Fig. 8)."

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.

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

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

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-month-long (?) period.

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.

Figure 11 caption is incomplete and too short.

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

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 $\sim$460 ppm. Is there an error here?

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

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.

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.

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

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

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.

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

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-183, 2017.

Printer-friendly version

Discussion paper