

Summary: While this draft shows improvement, more work is needed on the introduction/related work to set the stage for a strong paper. These sections should clearly set up: 1) what is already being done in the space; 2) what is lacking in the space; 3) what you will do differently to expand on what's already been done. There are plenty of other papers already using ML, GB., etc. – what about your model is different? Likewise, figures and tables should be included selectively – many are still superfluous and either demonstrate the same information as each other, or information that is already well-established in the field. These figures should be combined or removed as appropriate.

Subscripts are needed throughout for  $O_3$ ,  $CO_2$ , etc.

Abstract: Readers will know what ozone is. This space would be better spent explaining why you need machine learning enabled calibration.

Line 14: do the authors ever come back to these guidelines? If not, this paragraph is not useful. Same with the next paragraph – these standards are not really mentioned again later. I understand that this is trying to establish the “why use low-cost sensors”, but it needs to be more clearly related back to what you’re actually doing.

Line 27: Why is “primary” in quotes but not secondary? Be consistent, but quotes are not necessary. There are also quotes around primary on line 2.

Line 34: What is “official equipment”?

Line 51: While not detrimental, this paragraph is unnecessary.

Related works: see comments about table 15, but the information on the specific other sensors used for comparison could be restructured, if not removed. The information added here on specific ML models here is helpful, but it could be improved further by exploring more clearly the strengths and weaknesses of each of these, and how you will improve upon this and not just repeat what's already been done.

Table 3: This table should be removed. There are still no units in this table (temp, RH, PM2.5,  $CO_2$ ,  $NO_2$ , CO, etc. should all have units attached). The statistics of the measured quantities are not referenced or used anywhere else in the paper, and the reader can't do anything with this information on their own. Likewise, “stationarity” and “percentage of samples taking Different values” are not analyzed further in the text. The paragraph beginning on line 181 can be condensed to give the context the table is hoping to provide (ex. “Sensors X, Y, and Z appeared particularly unreliable and were omitted from our model”).

Figure 3: Any ambient pollutant will have a repeating diurnal pattern from the boundary layer rising and falling each day, and most sensors will pick up on major sources like traffic.

A DFT is not necessarily needed to show this and confuses the messaging in this section. Since this figure is never referenced again other than to show that a pattern exists, it should be omitted.

Figure 4: While this figure is fine, it's well known in the low-cost sensor space that sensors can capture the general trends of pollutants but need calibration to accurately convey the magnitude. This figure should be omitted.

Table 4: If this is all to make a better ozone model, the FIA of ozone should be included here to show how much it improves the model. How was 8% importance selected? It sounds arbitrary. It would also be easier on the reader if the threshold and the table were in the same format (either both in decimal or both in percent).

Figure 5 is essentially showing that some sensors are more cross sensitive than others, which is already well established in the field. This figure should be omitted.

Tables 5, 6, 7 and 8 should be combined into a single table with the 4 sub-categories as another column.

Table 9 is unnecessary and can be omitted – you and many others have already established that hyperparameter tuning will make the models fit better.

Tables 10 & 11 should also be combined.

For tables 9, 10, and 11, and Figure 6, it is not specified in the titles whether it is training or testing data – please specify.

In the low-cost sensor field, it is standard to show both training and testing data - consider adding to tables 9, 10, and 11, and Figure 6.

Is the point of Figure 7 just to show that the model isn't overfitting? It needs more analysis in the text rather than relying on the reader to interpret.

Again, it is well-established that low-cost sensors need calibration, and that tuning will improve models. Figure 8 should be omitted. If you are insistent on including something like this, an analysis showing the statistical significance in model improvement might be more impactful.

Is there more analysis or more takeaways to be had from Table 12? All the text is really saying is that the numbers in the table match the numbers in the figure. Stronger analysis in the text is needed to make the table worth keeping.

Is there a better way to visualize the information in table 13? It's inclusion is helpful, but a figure could be more informative than a table.

Table 14 contains repetitive information and should be removed.

Table 15 would be more useful if combined with table 1 instead of expecting the reader to remember the sensor specs from the very beginning. However, as the authors point out, this is comparing multiple different types of sensors that aren't inherently comparable. I understand that the authors are trying to show the usefulness of their calibration, but I don't think they need to directly compare with others for that message to come across. I recommend removing tables 1 and 15, especially because the inclusion of information on these other sensors in the earlier sections muddles the message of what the paper is ultimately trying to convey.

Line 321: This paragraph isn't indented, but all the others in this section are.

Line 324-325: Which model are these statistics from? The abstract suggests GB, but this should be clearly stated in the conclusions as well.

Line 350: Missing a period at the end of the sentence.