Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-129-RC1, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "Robust statistical calibration and characterization of portable low-cost air quality monitoring sensors to quantify real-time O_3 and NO_2 concentrations in diverse environments" *by* Ravi Sahu et al.

Anonymous Referee #1

Received and published: 16 July 2020

This manuscript describes calibration of low-cost electrochemical sensors for O3 and NO2. The sensors were co-located with reference monitors in Delhi and Mumbai. A novel part of this study was that some of the sensors were swapped between cities.

Overall, it is clear that the authors have put in a huge effort to build and test various calibration models. The effort to try different models seems truly comprehensive. However, they also seem to have tried to put everything into a single paper, and as a result it is very hard for me to follow the manuscript and to see how the conclusions follow from the evidence presented in the manuscript.

Printer-friendly version

Discussion paper

A revised manuscript needs to be much more focused. The current version tries to cover (1)different calibration algorithms (and somehow manages to both overwhelm the reader with lots of algorithmic detail but still put much - perhaps too much - information on some algorithms in the SI), (2) the swap experiments (which get lost in the discussion of the results), (3) the value of including T and RH in the calibrations, and (4) issues of data paucity and variability. In the end it is just too much to cover in one manuscript. The authors should think critically about the central message and try to communicate that clearly.

Specific comments: Abstract line 15 - it is unclear if the Rˆ2 improvement of 9% means 0.09 (e.g., 9 percentage points) or that the new Rˆ2 is 1.09 times the old Rˆ2. Also, if this is an important finding, it should be obvious in the manuscript text. I cannot point to the figure or table that directly supports this improvement.

Section 2.4 does not seem like it needs to be its own section. This information can be added to the end of section 2.3.

Line 203 - why are reference O3 concentrations <1 ppb scrubbed from the dataset? It seems entirely possible to have near-zero ozone in urban areas at night.

Section 3.3.1 describes splitting the data into 70% for training and 30% for testing. How is this different than the sub-sampled 2500 data points from section 3.3, item 2 in the list? (Also, it is not clear in section 6 when the 70:30 split was used versus the 2500 point subset.) The 70:30 split was repeated 10 times. I have seen 10-fold validation with a 90:10 split, which means that each data point appears in a testing dataset once. In your application of a ten-fold 70:30 split, it seems like there could be an uneven distribution of how often data show up in the training and testing subsets. Please explain why you used a 70:30 split with a 10-fold validation and give some context why this is a good approach.

Section 4 notes that "Our study used a large number of both parametric, and non-parametric calibration techniques." Those should be enumerated and briefly explained

in the text, rather than having readers hunt through the SI.

Section 5 proposes using k-NN without much explanation. Was some process used to select k-NN as the best algorithm or the one worth highlighting? It might be possible to move all of section 5 to the SI. I think the manuscript would benefit from getting to discussion of the results more quickly.

Figure 5 shows O3 increasing at site M in the middle of the night. This seems unrealistic at a time when the sun was not shining. The authors should explain what is happening or select a different example period to highlight.

The individual panels of Fig 6 should be labeled with the quantity shown. E.g., 6a should indicate NO2 on the figure panel, rather than having to dig through the caption.

Line 404 - what are RT and KRR? Those terms have not been introduced yet. General comment - this paper needs a glossary. It is very hard to keep up with all of the abbreviations.

I don't think that Figure 7 is appropriate for the main text. It should go into the SI.

Section 4 referred readers to the SI for details on the various calibration models. Now the models are described in section 6.2.2. This seems inefficient. Furthermore, the list of models should be comprehensive. Table 2 and Fig 8 both mention NW models, which of course are not listed in 6.2.2 and I had to go back 100 lines in the manuscript to find them!

General comment on length: The real meat of the paper starts in section 6 around line 425. However at this point I am pretty lost about the goals of the paper and what models are being compared. The paper is just much too long. I think most readers will quit before they get to the results.

I do not understand what Table 3 is telling me. Does interpreting this table require a good understanding of the Mahalanobis metric?

Figure 9 - please label the subpanels (a), (b), (c), (d) and explain each panel in the caption.

All of the violins in Fig 8 and 10 have instances of very high error. Are these occasions the same for all tested models? And what conditions lead to high error? I think that AMT readers will be more interested in model performance - e.g., periods when the models perform poorly - than the depth of mathematical detail presented e.g., in sections 4 and 5.

Section 6.4 seems like it should either be removed or significantly expanded. Right now it does not seem to add anything to the manuscript.

One of the main aspects of this study, at least as explained in the Intro, was the sensor swap between sites D and M. As far as I can tell, this was not discussed at all in the Results and Discussion. What happened to sensors moved from D to M and vice versa? Performance of calibration models built in city A and applied in city B is of major interest to the low-cost sensor community, especially as more sensors get deployed in locations that currently lack reference monitors needed to build local calibration models.

Minor and grammar Line 94 says that the sensors "include contain". Please edit.

Line 118 - does the Plantower output milligram/mˆ3 or microgram/mˆ3?

Line 133 - sites D and M are undefined.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-129, 2020.

Printer-friendly version

Discussion paper