Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-129-AC1, 2020

Atmospheric Measurement Techniques Discussions

Open Access

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.

Ravi Sahu et al.

snt@iitk.ac.in

Received and published: 6 September 2020

Response to Anonymous Referee #1

==============================

We thank the referee for the comments and suggestions. Below we offer some clarifications in response to the same. The revised manuscript will be communicated to the editorial desk shortly.

Printer-friendly version

Discussion paper

1. Focus and length of manuscript: A major revision to the manuscript has been prepared that offers a more focussed discussion. The revision reorganizes portions that significantly impact the overall discussion of characterizing calibration models. All other material e.g. Figure 7, has been relegated to the supplementary material. We thank the referee for pointing out the inefficiencies in the structuring of the paper.

2. Formatting: we thank the referee for taking pains to point out several improvements in typography and formatting e.g. merging subsections, adding a glossary, labels in Figures 6, 9 etc. We have incorporated all of them in the revised version of the manuscript.

3. Abstract: improvements are indeed in percentage points. In the revision, we have included tabular values in the supplementary material which explicitly quantify the improvements, in addition to the violin plots which offer a more visual interpretation. These values are referred to in the main text in the revised version.

4. O3 concentrations: we thank the referee for pointing this out. We indeed discarded only those reference monitor values that were less than 0ppb and have corrected this typographical error. The reference monitors sometimes offer negative readings when powering up and under some other anomalous operating conditions e.g. condensation at the inlet. However, we note that less than 0.1% of the valid timestamps had reference O3 values between 0 and 1 ppb.

5. Train-test splits: we chose a 70:30 split since it gave us sizeable sets for both training and testing. Machine learning and statistical estimation literature uses various splits such as 70:30, 80:20, etc. Our splits were repeated independently 10 times to allow two-sample tests to be carried out. The k-fold split method as mentioned by the referee, is another alternative. However, with k=10, the resulting 90:10 split offers a rather small test set which we wished to avoid. We verified that the choice of the size of the split (e.g. 70:30 vs 80:20) does not alter the conclusions of the paper.

6. Subsampled datasets: to create the subsampled datasets in section 3.3, we took a split (a split being a 70-30 division among train and test) and randomly subsampled

2500 points from the training portion of the split. The test portion was not altered since the aim of this experiment was to study how lack of training data affects calibration performance. To be more specific, if a dataset contained a total of 10000 valid timestamps, the train-test splits would contain resp. 7000-3000 points. For the subsampled version of this dataset, we would sample 2500 points from the 7000 points, train on those 2500 points and then test on the 3000 test points.

7. Motivation behind choice of k-NN: k-NN and kernel estimators (kernel ridge regression (KRR) and Nadaraya-Watson (NW)) are well studied non-parametric estimators in literature. These are also known to be asymptotically universal which theoretically guarantees their ability to accurately model complex patterns when given diverse and sufficient data.

8. Figures: we thank the referee for pointing out the improvements to the figures. We have replotted figures 5 and 9 in the revision to consistently show results across the same two full days of operation (01-02 July and 20-21 Oct) for sake of clarity. These figures in the old version chose different days as well as different durations which we agree was inefficient. Figure 5 had a manual labelling error (high O3 levels in the night) which we have corrected in the revision.

9. Table 3: A metric tells us how to compute distance between two points, say 8 dimensional vectors in our case. The Euclidean metric gives equal importance to all 8 dimensions when calculating distances. An alternative interpretation of a Mahalanobis metric is that it tells us how to reorganize dimensions/features so that the resulting distances, when used by the kNN algorithm, give better performance. Table 3 shows us the optimal reorganization found by the metric learning technique. In particular, note that it places heavy emphasis on the Rh and T features. This means that the optimal Mahalanobis metric identifies that a high importance should be placed on Rh and T features when computing distances for use by kNN.

10. Regions of high error: The (revised) supplementary material now contains an

Printer-friendly version

Discussion paper

analysis of regimes in which various algorithms offer larger errors. The (Rh, T) space was divided into various buckets to analyze the performance of each algorithm in each bucket. For data hungry non-parametric algorithms such as RT, NW(ML), KRR, and KNN-D(ML), regions of larger error coincided almost entirely with regions where data was scarce. This is as expected. The least squares method on the other hand demonstrated no such clear trend on regions of high error. We also tracked the errors of various algorithms across the day and found that for O3, whose diurnal levels are more predictable, all algorithms tended to offer relatively larger errors when the (true) O3 levels were higher (i.e. during peak sunlight hours). For NO2, which demonstrates no such predictable diurnal patterns, no patterns in errors were observed either.

11. Section 6.4: we submit that sections 6.5 through 6.8 were meant to be subsections of section 6.4. We have corrected this formatting error in the revision.

12. Swap experiment: Table 5 does discuss cases when sensors are trained in one season and tested in another season. Cases are considered when the site is kept the same across seasons, as well as when the site is changed across seasons. We also request the referee to take a look at the comment of Referee #3 on this point and our rebuttal to the comment (please see “General Comments” bullet point 2 in our response to Referee #3).

13. The Plantower PMS7003 offers readings in microgram per cubic meter. We thank the referee for pointing out the correction and have made the same in the revision.

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