Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-216-RC2, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "Development of a General Calibration Model and Long-Term Performance Evaluation of Low-Cost Sensors for Air Pollutant Gas Monitoring" *by* Carl Malings et al.

Anonymous Referee #2

Received and published: 5 November 2018

General comments: In this study, authors compared different models used to adjust data from low-cost sensors and compared model application in different sites within the Pittsburgh, Pennsylvania region of the USA. These comparisons demonstrate that at the sites different model-based calibration techniques show differing abilities to accurately produce concentration data for the various sensors employed. However, there existing two main problems: 1 for calibration period and deployment, it is unclear which based on the text, the calibration data also used for testing or vice versa, more details about the deployment should be illustrated. 2 Evaluation validity: As listed in Table 4, the average pollutant concentration is relatively small for NO and NO2 both in training and testing period, which about 1.7 and 6.4 ppb level, how did the investigators ensure

Printer-friendly version

Discussion paper

sensor evaluation validity in such low-level situation, while EC sensor has a lower detection limit at about this level? And for NO, the MAE of all models are even large than average concentration.

Specific Comments: 1. Page 1, Line 24: Authors use “These results will help guide future efforts in the calibration and use of low-cost sensor systems worldwide”. Perhaps the approaches merit many further trials in the widely differing pollutant and meteorological conditions It is unclear. It is unclear how the study demonstrates protocols that apply worldwide?

2. Page 2, line 16: “These sensors tend to have lower signal-to-noise ratios than regulatory-grade instruments. . .” is a fairly unusual and non-specific way to say they are not as precise or sensitive as conventional air monitors.

3. Page 2, Line 20: Authors mentioned that a nonlinear interaction in the reference (Jiao et al., 2016), however, this paper didn’t mention this kind of nonlinear interaction, but cited Gao’s paper (Gao et al., 2015) which also didn’t mention this, rather it is a nonlinear response for PM sensor.

4. Page 2, Line 25: Same problem for reference (Cross et al., 2017) as comment 3.

5. Page 3, line 29: It appears that on-campus field calibrations were conducted during a brief period in summer and early fall. Where winter time calibrations performed as well? If not, how might this impact the application of various results of modelling over winter time conditions?

6. At the end Section 2. which describes monitoring methods there is a general reference to the Zimmerman 2018 paper to provide the reader information on protocols. Key factors should be pulled into the current text or figures. Specifically, on Line 30âĂŤwhat were the specific models of regulatory monitors used and how were they located/protected from environmental factors? What methods were employed to measure speciated VOCs mentioned? What is meant by the term “or BTEX”? This does

not appear to be covered in the 2018 Zimmerman paper and it does not seem that VOCs are part of this studyâĂŤperhaps this needs to be removed from the current text. Further, SO2 is included in the list of pollutants that appear to have been included in this study. However, no data for SO2 appears. Further, several of the models include factors for SO2. Was SO2 measured or employ in this study?

7. Page 4, Line 5: The collocation in ACHD points seems to span the whole year of 2017 in each week at Lawrenceville as shows in Figure 7, while in Table 4 the maximum days of testing period duration only ranged from 75 to 110 days for CO, O3 and NO2. The authors should give more explanation about the deployment? It is unclear and could not be found in the cited reference (Zimmerman et al., 2018). Also, as mentioned in Page 5, Line 9, 80% of collocation is about 28 days less than 1 month, what are other days while RAMP collocation such long time and what is the criteria for period selection?

8. Page 4, Line 10: Authors mentioned ACHD Parkway East has high levels of NO and NO2 for maximum at 100ppb and 40ppb, this period also included in testing period, but in Table 4, for whole testing period these two maximum concentrations are 66 and 32.

9. Page 4, Line 22: The best model was selected based on performance of individual models, what is the criteria for this?

10. Page 9, Line 4: The random forest and hybrid model used all sensor data for training CO, and combined linear model which only used CO, why not introduce other sensors in this procedure?

11. Page 10, Line 8: As mentioned in the text “measurements where the corresponding true value is below an assigned lower limit are removed from the measurement set to be evaluated”, while for NO2(10ppb) and O3 (10ppb) which listed in Table 1, the training period or testing period concentration range still start from 0 or 1ppb, does any filter about lower limit come into force in the training or evaluation? The authors should discuss the impact of removal of data below specified minimums as this would

appear to skew the actual data and do so in a differential basis depending on ambient conditions.

12. Page 16, line 11: “The fact that for most gases a variety of calibration approaches show similar and (for typical uses cases, acceptable) performance may reflect better underlying performance from the RAMP monitor, as similar studies for other low-cost sensor packages showed a wider variability in performance between calibration approaches (see e.g. the summary provided by Zimmerman et al., 2018).” This statement appears to indicate that the RAMP monitors somehow performed better than other similar sensor based monitors. However, there is no rationale for this statement based on the actual monitor package, component details provided or information reported by others. It appears to simply be a diffusion based sensor deployment of Alfasense electrochemical cells and an NDIR CO2 sensor. What other factors might influence system performance?

13. Page 25, Figure 4: The performance of CMU site for each sensor is marked in hollow marker, while there are two hollow markers in CO and NO2, which is unclear.

14. Page 30, Table 4: In the evaluation of testing period, R2 was used, however, in previous figure and text, such as Figure 2-4, Pearson linear correlation coefficient r was used, what is the standard for selection between these two parameters?

15. Page 30, Table 4: The title of the paper is about a general calibration model (gRAMP) and recommended in the conclusion, while it was not compared with iRAMP models in Table 4.

This paper focuses on field calibration methods and model application to produce adjusted pollutant concentration data from sensor based monitors. However, discussions and recommendations are only made regarding field calibration approaches. It would also appear that the data agreement between sensors and regulatory monitoring might be improved by other means. For example, conventional air monitoring methods as applied to regulatory monitoring efforts always include periodic zero and span challenges.

Improved circuitry or air conditioning might aslo improve monitor performance. How might the inclusion of such methods improve the quality of data produced by sensor based systems?

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-216, 2018.

Printer-friendly version

Discussion paper