Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-260-RC1, 2017


Atmospheric Measurement Techniques Discussions

# *Interactive comment on* "Closing the gap on lower cost air quality monitoring: machine learning calibration models to improve low-cost sensor performance" *by* Naomi Zimmerman et al.

**Anonymous Referee #1**

Received and published: 11 September 2017

This work presents calibration models for low-cost sensors of NO2, O3, CO, and CO2 that show encouraging results toward use of such devices in exposure analysis and dense measurement networks. The authors describe the instruments, calibration algorithm, and evaluation metrics in a pedagogical way that is easy to follow. The most advanced algorithm (Random Forest) performed the best, and from it the authors were able to extract the importance of each variable to the calibration to provide better understanding of these statistical models. The authors place the performance of their new calibrations in context of simpler calibration models, previous studies on sensor calibration, and several performance guidelines established by regulatory agencies. The machine learning approach appears to take into account interfering cross-sensitivities



to other pollutants, but at times predict by correlation (e.g., for CO2 and NO2) rather than direct signal response to the pollutant. The manuscript is well-written and explains an approach that allows identification of important explanatory variables for a statistical calibration model, shows that several calibration models are able to maintain reasonable performance even in lower concentrations than demonstrated in previous studies. The manuscript presents a relevant contribution to the emerging literature on low-cost sensor calibration and is therefore recommended for publication in Atmospheric Measurement Techniques after the following comments have been addressed.

General comments:

The title is a bit ambitious, ambiguous, or both. How much of the performance "gap" is closed by a) improved hardware compared to past studies, b) the algorithm (i.e., Random Forest), c) sensor combinations at each node, and d) range of different sample types collected? Application of machine learning for sensor calibration in the field has been performed before, but the title and abstract seems to give the impression that this reduces the gap. There is much focus given to RF but there is no indication that it has an inherent advantage over other machine learning methods. For instance, it is possible that a MLR model could also handle cross-sensitivities only if it were provided all variables (though RF and other machine learning algorithms are more flexible in that it does not require the assumption regarding global linearity).

The past work of De Vito et al. (2008, 2009) also show encouraging results from a long-term evaluation of field calibrations (for low-cost multi-sensor devices for benzene, CO, and NO2 against government monitoring station instruments using machine learning algorithms).:

De Vito S., Massera E., Piga M., Martinotto L., and Di Francia G.: On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sensors and Actuators B: Chemical, 129(2):750–757, doi:10.1016/j.snb.2007.09.060, 2008.

De Vito S., Piga M., Martinotto L., and Di Francia G.: CO, NO2 and NOx urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization, Sensors and Actuators B: Chemical, 143(1):182–191, doi:10.1016/j.snb.2009.08.041, 2009.

The manuscript is perhaps too bold in its tone. Accurate predictions are shown for concentration (and T, RH) domains that are present at the location of the reference monitor used for calibration, even while using different data points. (As stated by the authors, current implementation of RF is limited to the domain of the training set.) Dense network coverage implies monitor placement in different microenvironments (e.g., near-roadway, etc.) which would experience different concentration regimes. Moreover, some of the explanatory variables used for calibration may be surrogates for another variable which may vary differently at another site. There is mention of two RAMPS units deployed in Pittsburgh and their positive evaluation against other reference measurements in a mobile van (p. 17, line 15), but no results are shown.

Since corrections of the supersite reference monitors against the Allegheny County Health Department instruments are necessary, why not make this Allegheny County Health Department site the reference site? Given the local contributions of vehicle emissions to CO and NO2 that are present in the parking lot site, how were the corrections for baseline drift determined?

While the authors describe the use of 5-fold CV to selection the explanatory variables to use, the choice of 5 data points per terminal node / 100 trees per fold does not seem to be explained. This was also selected in the CV process?

p. 14 Line 18 paragraph: Is this not possibly a limitation of the hardware?

Minor comments:

Section 2.2: Data coverage (i.e., missing data) and the time resolution should be stated here rather than (or in addition to) later in the manuscript.



P. 9 Line 15 to end of paragraph. The authors switch from describing "intermittent" collocation to "distributed" collocation. Given the discussion of multiple RAMP monitors, "distributed" can be confusing. Also, "degree of collocation" is referring to frequency or effective duration?

p. 10 Line 19: value of correlation for NO2 and CO2 with reference monitors is missing.

p. 10 Line 22: insert figure numbers (SI Fig S3-S6).

p. 10 Line 30: The relationship between m_try and model complexity is not very clear.

p. 10 Line 13: "clearly outperformed" -> not for CO

p. 11 Line 21: insert figure numbers (SI Figs S7-S10). Slopes, correlations, or some of the metrics listed in Table S2 included in the panels would be informative. Why are some RAMPS not included?

p. 11 Line 31: "NO2" -> "O3" here?

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-260, 2017.