Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-473-RC3, 2021

Atmospheric Measurement Techniques Discussions

Interactive comment on "Towards low-cost and high-performance air pollution measurements using machine learning calibration techniques" *by* Peer Nowack et al.

Anonymous Referee #2

Received and published: 17 May 2021

This study investigates a promising calibration method for the low-cost air pollution sensors, through co-location with public measurement stations and regression towards the station reference data. Four machine learning algorithms for the regression, namely Ridge, Random Forest (RF), Gaussian Process Regression (GPR) and Multiple Linear Regression (MLR) are implemented and compared. The influence from co-location training conditions and choice of calibration input data, and the issues of "site-transferability" are also discussed. Overall, this interesting manuscript is well written, and potentially contributes to the scientific community and also the environmental protection agencies. My comments are listed below:

1. My biggest concern is about the "site-transferability". 1) Any more insights on the key factors limiting the site-transferability? Description between Line 102-110 has stated that the PMS5003T series PM particle sensors are based on laser scattering using Mie theory, and the composition-wise similar environments are assumed for measurement and calibration, which potentially limits the site-transferability. If the authors could also explain the measurement principles for other sensors and how would they potentially limit the site-transferability? 2) The use of multi-sensor node is a very good idea to involve more factors for the regression and thus potentially enhance the "site-transferability". But the description for the set-up of sensor hardware in Sect. 2 is not clear. In Line 99, "Each multi-sensor node contained", does it mean that all nodes listed in Table 1 have the same set-up of sensor hardware? If so, why the same set-up is used for NO2 and PM10, respectively? In Table 1, different nodes spanning over different measurement time? Why so many nodes are used at CR7 site for the relatively shorter time period? Please add clarification. 3) In Figs. 7-8, the results for "site-transferability" looks promising except NO2 concentrations predicted at CR9 between $\sim$700-1000 h are largely underestimated. What if the node-19 is calibrated at CR9, would the peak concentrations between $\sim$700-1000 h be properly predicted? Why the calibration is not performed at the site with a larger range for the observed concentrations (e.g., CR9 rather than CR7 for NO2 for the Sect. 3.1)? I mean, the calibration at CR7 discussed at Sect. 3.1 seems to have a limited applicability.

2. Besides the spatial variability, the temporal variability is also very important. I mean, given the fixed co-location, if the calibration performed during the winter time, is applicable to make proper predictions during the summer time? Furthermore, any special treatment for the calibration of the sensor signals during occasionary condition (e.g., intensive washout during the heavy rain, dust event, or strong temperature inversion, etc.) when the observed concentrations at the reference site are extraordinarily high or low?

3. In Table 2, it is very interesting that the performance is greatly enhanced if the sensor

signals relevant with NO/O3 are involved into regression (i.e., (g) v.s. (f)), and further involvement of more factors does not seem to improve the performance significantly. Any more detailed explanation (the description between Line 375-378 lacks of insights)? How about the performance for the NO/O3/T/RH, if it is better than A43F/T/RH and B43F/T/RH? And the computing cost for each selection of predictors?

4. It would be very nice if the authors could use a table to summarize and compare the key assumptions, algorithms, advantages/disadvantages, computing cost etc. for the four regression methods.

5. In Fig. 5, how the circles (for each individual node) are calculated, as each node has a different time span? Table 2 and Fig. 4 are based on the average of the 21 nodes, right? Please add clarification.

6. Why PM10 rather than PM2.5 is focused on? PM2.5 is more relevant with the public health issues.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-473, 2020.