Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., author comment AC1
https://doi.org/10.5194/amt-2022-65-AC1, 2022

Reply on RC1

Priyanka deSouza et al.

Author comment on "Calibrating networks of low-cost air quality sensors" by Priyanka deSouza et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-65-AC1, 2022

Reviewer 1:

This is a very timely paper that provides a systematic and deep analysis on the different ways that low cost sensors can be calibrated by colocation with regulatory grade equipment. In particular, it provides useful information on how best calibrate depending on the colocation period possible. The paper uses a variety of calibration models (n=21) starting with simple linear corrections, and ending with complex machine learning algorithms, where it is often difficult to know the mechanism of the correction. The calibration models are tested on four different colocation periods. In particular the difference between the C1 and C2 colocation strategies is interesting because it shows that more calibration data is not necessarily helpful if it doesn't capture the variability in the parameters. The hot spot analysis is also interesting, highlighting the need for care when interpreting individual sensors within a network.

Low cost sensors are used in various ways. Sensor networks like the 'love my air' network used as the data set in this paper are used to complement existing regulatory activities, whereas in other contexts low cost sensors are used where regulatory measurements are scant or non-existent. This paper will provide very useful to all users of low cost sensors.

The paper is very robust in its description and should be published, once the following (mostly minor) points are addressed.

Authors: Many thanks for this assessment of our work

In general, the resolution of the figures should be improved.

Authors: Thank you. We have improved their resolution in this revision.

Abstact and L49 – no need to say 'gold standard reference monitors', 'reference monitors' is sufficient.

Authors: Thank you. We have removed the extraneous term

L42 estimates vary widely for number of premature deaths due to air pollution, this should be acknowledged, or at least the prefix of 'approximately' should be added by the 6.7M.

Authors: Thank you. We have included the pre-fix 'approximately'

L70 'leading to mass overestimation…' should be 'leading to the (regulatory) dry mass overestimation' or similar

Authors: Thank you. We have noted that we mean dry mass estimation

L74 need to acknowledge that most of the PM mass concentration is at particle diameters greater than 300 nm.

Authors: Thank you we have done so

"LCS are not able to detect particles with diameters below a specific size, which is determined by the wavelength of laser light within each device, and is generally in the vicinity of 0.3 μm, whereas the peak in pollution particle number size distribution is typically smaller than 0.3 μm."

L96 Köhler not kohler

Authors: Thank you. We have made this change

L119 I would state that R^2 is a misleading indicator rather than might be

Authors: Thank you. We have made the suggested change

L215-216 you would expect averaged data to have less variance.

Authors: The necessary scale of the plots (to capture spikes of minute-level $PM_{2.5}$ as high as ~1000 μg/m^3) which perhaps make it harder to evaluate variability. When we zoom into a smaller subset of sensors as in Figures S4 we see a high degree of variability

L240 RH, T, and D are not independent parameters. A discussion of the use of non-independent parameters within the calibration algorithms should be provided.

Authors: The reviewer is quite right. We note, the following when describing D:

"We derived dew-point (D) from T and RH reported by the Love My Air sensors using the *weathermetrics* package in the programming language R (Anderson and Peng, 2012), as D has been shown to be a good proxy of particle hygroscopic growth in previous research (Barkjohn et al., 2021; Clements et al., 2017; Malings et al., 2020). Some previous work has also used a nonlinear correction for RH in the form of $RH^2/(1-RH)$, that we also calculated for this study."

We note that we use D, in addition to T and RH because for all our multilinear regressions we used the same set of equations that US EPA researchers used when deriving a national calibration equation for PurpleAir monitors.

To whit in our note on statistical modeling we add the following note:

"Sixteen models were multivariate models that were used in a recent paper (Barkjohn et al., 2021) to calibrate another network of low-cost sensors: the PurpleAir, that rely on the same $PM_{2.5}$ sensor (Plantower) as the Canary-S monitors in this study. As T, RH and D are not independent (**Figure S8**), the 16 linear regression models include adding the meteorological conditions considered as interaction terms, instead of additive terms. The remaining 5 relied on machine learning techniques."

L302 how do you choose which site to leave out in the LOSO methodology? What potential bias(es) does this introduce into the analysis?

Authors: Thank you for this question. We left out each site in turn and used models developed for the other sites to make predictions at the left out site. We chose the model that yielded the best average performance across each of the left out sites. We include this description in the text:

"We used a Leave-One-Site (I25 Globeville, I25 Denver, La Casa, CAMP)-Out (LOSO) approach for cross validation (CV) to prevent overfitting in our machine learning models (Models 17 - 21 in **Table 2**). Briefly, we split the data into four groups, with each group excluding data from a single reference monitoring site. In each cross-validation iteration, we selected each group in turn to fit the model and made predictions at the left-out site. **The model that had the best average performance across all the left out sites was chosen. In this manner this CV approach was used to tune the hyper parameters in the machine learning models adopted in this study using correction approaches: C1, C2, C3 and C4.**"

L333 and most other equations. Pet peeve – use proper multiply symbol rather than x in equations.

Authors: Thanks. We have made this change everywhere.

L351 "as these concentrations account for the greatest differences in health and air pollution avoidance behavior impacts" this statement is unclear. Are you suggesting that 30 ug/m3 is a cut off for more harmful PM health effects? My understanding is the health effect: concentration curve is reasonably linear over these ranges.

Authors: Thank you for this note. Our choice of this threshold is derived from the way in which AQ Health Index (AQHI) is derived. More information on this threshold can be found in the paper we cited: Nilson et al., (2022)

L393 note a p value of 0.05 means that 1/20 results can be to chance. With 21 models and 4 colocation conditions, you might expect some false positives.

Authors: Thanks for this note. In the context of this paper, we compared the distribution of errors in prediction on each test dataset when leaving out a given site to errors derived from using data at all co-location sites. A $p < 0.05$, indicates that for each comparison, there is a 95% probability that the errors belong to the same distribution. This threshold is widely used.

L457 model 2 has a lower RMSE than model 16, so doesn't that contradict "more complex models yielded a better performance"

Authors: Thank you for catching this mistake. For Model 2, instead of listing the RMSE, we made a mistake and listed R instead. We have corrected this as follows:

"We also found that for corrections C3 and C4, more complex models yielded a better performance (for example the RMSE for Model 16: 2.813 $\mu g/m^3$, **RMSE for Model 2: 3.110 $\mu g/m^3$** generated using the C3 correction) when evaluated during the period of co-location, alone (**Tables S2** and **S3**)."

L472 "the nonlinear correction for RH" gave best performance. Doesn't this suggest a model using a physically reasonable model (essentially k-Köhler) works best when extensive colocation data is not possible. See for example Crilley et al. (2020) https://doi.org/10.5194/amt-13-1181-2020

Authors: Thanks! We indeed make this observation in the Discussion and think it is an important take-away from this paper:

"For C3 and C4, we found models that relied on nonlinear formulations of RH, that serve as proxies for hygroscopic growth, yielded the best performance, as compared to more complex models. This suggests that physics-based calibrations are potentially an alternative approach when relying on short co-location periods and need to be explored further."

L528 does the temperature offset on CS19 make sense with respect to the position of the sensor?

Authors: Yes it does. It is in the shade.

Please also note the supplement to this comment:
https://amt.copernicus.org/preprints/amt-2022-65/amt-2022-65-AC1-supplement.pdf