Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., author comment AC3
https://doi.org/10.5194/amt-2021-425-AC3, 2022

Reply on RC3

Brayden Nilson et al.

Author comment on "Development and evaluation of correction models for a low-cost fine particulate matter monitor" by Brayden Nilson et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-425-AC3, 2022

Thank you for the time you put into reviewing our manuscript and the very useful and helpful feedback which has led to improvements in the paper. Please see our following responses and proposed alterations which we believe will resolve your individual comments.

Comment 1:

Line 123: Please state the reasoning behind your choice of 5ug/m^3 as the absolute error cut-off for identifying failures in either sensor. Is this a recommendation from PA?

Response 1:

We added a statement that this method and cutoff was derived from methods proposed by Barkjohn et al. 2021 and Tryner et al. 2020.

Comment 2:

Line 138: "The final set of colocation sites (47 in total) were then selected as those with at least half a year (4380 hours) of valid data from both PA and FEM and a minimum correlation of 50% for all valid hourly observations over the period of record." What is the reasoning for setting the minimum correlation to 50%? While this would remove any non-collocated sensors wouldn't this also possibly remove any poorly performing collocated sensors?

Response 2:

The minimum correlation of 50% was iteratively decided such that the sites that were clearly poor performers (i.e. those potentially not colocated) were removed, while known colocation sites were not. This method is not perfect, but a cutoff needed to be decided given the automated nature of our approach.

Comment 3:

Line 148: "A temperature term was also tested; however, its impact was found to be

minimal." Does this statement refer to the pooled together training dataset? Was this ever tested for individual sites?

Response 3:

Anytime we added temperature to a model it performed worse than a model with just RH or both combined (in general and at a few sites we initially worked with on the individual level). For clarification, in the revised paper we added "and given the high correlation between temp and rh, rh was selected as the more important term" to this sentence.

Comment 4:

Line 159: In equation 3 please clarify that the correction factors of a and b for PM2.5 < x are different than the factors a and b for x< PM2.5< x2.

Response 4:

We modified the table to reflect this.

Comment 5:

Line 191: How were the training and testing sites divided up? Rather than dividing by site did you consider dividing by time (randomly dividing to ensure similar conditions between testing and training datasets)? Please clarify the reasoning between choosing 32 training sites and 15 testing sites (~2/3 training).

Response 5:

We added the following to this section: "Training /testing sites were randomly selected then adjusted (again randomly) to ensure representativeness across geographic areas and concentration ranges."

Yes, we considered dividing by time, but given the seasonal differences that can occur (and the variation in this between locations) we felt this would either result in 1) one of the datasets (testing or training) having a disproportional amount of episodic concentrations (winter inversions, wildfire smoke, etc) or 2) the two datasets being highly dependent on each other (if you took every third day as test for example).

We initially had more testing sites than training sites, but we found that the other way (what we presented here) was more common in the studies we cited that had a large number of colocation sites.

Comment 6:

Line 246: "Further comparisons were only made on Models 1, 2, 7, and 8 as they showed the best performance here." It would be informative to include the results from the other models in a supplementary information section.

Response 6:

We added two figures and a table to the SI document with the results from the remaining models and added a reference to them at the start of the results section.

Comment 7:

Line 294: "The concentrations of PM2.5 reported from the PA monitors were biased high

compared to the FEM monitors at most colocation sites, especially for the lower concentration range." While you evaluated model bias at different PM2.5 concentrations did you consider looking at the bias over a RH range?

Response 7:

At extreme RH values the correction models can have an exponentially increasing impact on the raw data. We plotted all of our concentration data in a scatter (PA-FEM error on the y, RH on the x) and at RH outside of 30-70% the bias in PA changed noticeably (see new figure in SI).

Comment 8:

Line 324: For scenarios where testing models on individual locations is not an option, such as applying a correction in an area without a nearby PA-FEM colocation site, we recommend using our Model 2." Rather than use a model that has been trained on a variety of locations/conditions and therefore is pretty generalized, would it not be more prudent to use a model that has been trained on conditions similar to those you expect to encounter?

Response 8:

Yes, but this statement is for the scenario where those data are not available. At the moment there are many correction formulae that have been developed, several of our tested ones come directly from the PurpleAir map and do not perform as well. We posit that our Model 2 would be better to use than these other models for general application across many locations (like you would on a mapping platform).