

Interactive comment on “Development and Application of a United States wide correction for PM_{2.5} data collected with the PurpleAir sensor” by Karoline K. Barkjohn et al.

Anonymous Referee #2

Received and published: 27 December 2020

GENERAL COMMENTS

The authors present carefully-considered, practical guidance for performing quality assurance checks on PurpleAir data. The authors' model for correcting PurpleAir PM_{2.5} data benefits from a large training dataset that spans many months and locations (39 sites in 16 U.S. states). Together, the quality assurance guidelines and correction model presented here can help users improve the accuracy of PM_{2.5} data reported by PurpleAirs. There will likely be applications in which this U.S.-wide correction still does not produce sufficiently-accurate estimates of local ambient PM_{2.5} concentrations, but I am not here to argue for letting the perfect be the enemy of the good.

C1

SPECIFIC COMMENTS

1. Lines 79–82: I found these two sentences confusing. Are the PurpleAirs that are part of the “larger project” counted as PurpleAirs sent out by EPA (group 1 on line 77) or PurpleAirs independently operated by air monitoring agencies (group 2 on line 77)? Can the authors clarify how these PurpleAirs fit in? Alternatively, can the authors remove these sentences and simply describe the study as involving PurpleAirs “sent out by EPA to capture a wide range of regions and meteorological conditions” and “other collocated sensors”?
2. Lines 91–92: The authors state here that their analysis included data from 50 PurpleAirs at 39 sites in 16 states. On lines 87–88, the authors state that their search identified 42 sites in 14 states. Does that mean the 50 PurpleAirs/39 sites/16 states includes both PurpleAirs sent out by EPA and PurpleAirs identified in the search?
3. Lines 121–127: Was the “within 5 $\mu\text{g m}^{-3}$ or 61%” check performed on the 2-minute/80-second averages or on the 24-hour averages? Was this check performed on the cf_1 data or on the cf_atm data? Please clarify.
4. Section 2: Can the authors include some information on the breakdown of regulatory monitors and reference PM_{2.5} measurements by type (FRM vs. FEM and type of FEM)? Maybe this information could be included in Tables S1–S2 and/or summarized in the text? FEM measurements do not always agree with FRM measurements and some FEMs might be more likely to deviate from FRMs than others. I downloaded the daily PM_{2.5} data from the sites listed in Table S1, over the date ranges listed in Table S2, from the AQS. It looks like there were 17 FRMs, 16 BAMs, 16 spectrometers (T640s), and 6 TEOMs across the 39 sites. It looks like the breakdown of 24-hour measurements was approximately 10% FRM, 50% BAM, 30% spectrometer, and 10% TEOM.
5. Lines 153–154: It looks like there were 13 sites with both FRM and FEM data. It looks like there were three sites (53-033-0057, 53-061-1007, and 55-087-0009) with two FEM monitors and one site (19-113-0040) with two FEMs plus an FRM. What did

C2

the authors do when multiple reference measurements (FRM or FEM) were available for a single 24-hour period at a single site? Were all 24-hour average measurements included as separate data points? Or was just one selected and, if so, which one?

6. Lines 154–155 and 470–474: Did the authors look at the agreement between 24-hour average FRM and FEM measurements for the dates and sites where both were available? It might be useful for comparison to the agreement between PA and FM data. For the sites and dates listed in Tables S1–S2, it looks like the RMSE for 24-hour FEM measurements, compared to FRM measurements, was $1.7 \mu\text{g m}^{-3}$ for BAMs ($n = 4134$), $2.1 \mu\text{g m}^{-3}$ for spectrometers ($n = 4405$), and $1.1 \mu\text{g m}^{-3}$ for TEOMs ($n = 69$). Does a RMSE of $2.1 \mu\text{g m}^{-3}$ constitute “adequate accuracy”?

7. Line 149: How come the PurpleAir data had to cover 90% of the 24-hour period to be considered “complete,” but the FEM data only had to cover 75% of the 24-hour period?

8. Line 166: Why was dewpoint considered in addition to temperature and RH?

9. Lines 168–177: I agree that the model inputs should all be data from the Purple Air. I’m also glad the authors point out that temperature and RH data reported by PurpleAirs are biased high and low, respectively. I don’t agree with Reviewer 1 that temperature and RH values reported by PurpleAirs shouldn’t be used just because they’re inaccurate. Your model does not assume they’re accurate; it just assumes that they’re correlated with the actual temperature and RH (see Figure S6 in Magi et al., 2020 and Figures S12–S13 in Tryner et al., 2020). I’m not sure I agree that the temperature and RH reported by the BME280 reflect what the aerosol experiences during measurement. Inaccuracies in the PurpleAir temperature and RH data arise because there is no convective or forced airflow over the BME280 to dissipate heat from the breakout board or other electronic components inside the housing. The inlets to the PMS5003 sensors are at the bottom of the housing, and there is active airflow through the PMS5003 sensors.

C3

10. I agree with Reviewer 3 that correlation between input variables warrants further consideration. Can the authors add a correlogram to the supplement to indicate the extent to which different variables were correlated (lines 179–182)? Might the sensitivity of the model coefficients that the authors discuss on lines 347–349 be due to correlation between the independent variables? Are temperature and RH correlated? Could that be why temperature does not reduce error and bias in the nationwide dataset (line 362)—because it essentially provides the same information as RH?

11. Lines 202–206: Were these analyses performed using the model fit to the full dataset (Equation 8)? Or were they performed using the cross-validation methods described on lines 187–195? If the model fit to the full dataset was used, how do the authors justify that choice? Wouldn’t the model be expected to almost always predict the AQI correctly for the dataset it was fit to?

12. Lines 220–231: It’s great to see this thorough discussion of quality assurance criteria!

13. Line 239: Were these summary statistics calculated before reducing the size of the Iowa dataset and, if so, did any of these values change after?

14. Line 242: By “10,907 days,” do the authors mean 10,907 pairs of PA/FM 24-hour $\text{PM}_{2.5}$ concentrations? Were multiple data points from a single site on a single day (either multiple PurpleAirs or multiple regulatory monitors) treated as one day or as multiple days? Please clarify. This comment also applies to Table 1. There certainly weren’t 3762 days between 11/29/2017 and 12/29/2019.

15. Lines 310–312: I agree that it’s more appropriate to use the cf_1 data in your model instead of the cf_atm data displayed on the PurpleAir website. The cf_1 data explained more variance than cf_atm data and cf_atm data are known to be nonlinear at concentrations above $25 \mu\text{g m}^{-3}$.

16. Line 335: Does the “linear correction” refer to Equation 3?

C4

17. Lines 335–337: Is there a typo here? The text says that MBE drops from 3.3 to 0 $\mu\text{g m}^{-3}$ for LOBD and from 4.2 to 0 $\mu\text{g m}^{-3}$ for LOSO, but in Figure 4 it looks like the opposite: MBE drops from 4.2 to 0 $\mu\text{g m}^{-3}$ for LOBD and from 3.3 to 0 $\mu\text{g m}^{-3}$ for LOSO.

18. Sections 3.4 and 3.5: Did the authors look at whether/how their results varied if only FRM measurements were used as the reference or if only measurements from a single type of FEM (BAM, spectrometer, TEOM) were used as the reference? It seems like the authors' large dataset could be used to investigate this question, which might inform others who try to correct PM sensors by collocating them with regulatory monitors. Groups conducting smaller collocation studies that span a small geographic area might only have access to a single type of FEM, so it would be good to know how that might affect their results.

19. Lines 443–449: Do the authors think the lack of an RH term in the AQ&U model is the reason it overestimates $\text{PM}_{2.5}$ concentrations for the authors' dataset? RH is typically low in Salt Lake City, while the authors of this study have data from several sites in more humid locations.

20. Lines 489–490: Is the BME280 sensor inside the PurpleAir expected to fail before the PMS5003 sensors? Is there evidence (in the literature or from the authors' experience) that this has occurred in the past? If not, why is this concern emphasized?

21. Lines 512–515: Did the authors provide guidance to the air monitoring agencies on how to appropriately site the PurpleAirs? If so, can the authors relay that guidance in Section 2.1? It might be helpful to PurpleAir users. If not, would the authors like to comment on what they would consider ideal siting and/or provide examples of siting pitfalls? I feel like this might be appropriate given the emphasis on quality assurance elsewhere in the manuscript.

22. Line 530: On lines 150–151, the authors note that they included exceptional event days in their analysis. Did/can the authors identify specific events like dust storms and

C5

wildfires with the help of site operators? Have the authors considered looking at the performance of their model specifically during extreme events? I think the results would be of interest to readers.

23. Table 2: Do these metrics describe the performance of Equation 8? Equation 8 was fit to the data described by the "AB, completeness, problem sensors" criteria, correct? If so, I'm not sure it's appropriate to evaluate a model using the same data used to fit the model and then compare that performance to other datasets that the model was not fit to (rows 1–6).

24. Figure 4: I'm not sure that the ALL column adds useful information to this figure (for the reason given on lines 341–342).

ORGANIZATION

A lot of material is presented in a way that seems disjointed. One way to remedy this issue would be to consistently describe methods in the methods section and results in the results section. Another option is to organize the sections after the introduction as Data Collection, Quality Assurance, Model Development, and Model Evaluation instead of using conventional Methods and Results sections.

1. Lines 102–109: It's great that the authors (i) point out the need to look for erroneous temperature and relative humidity values during the quality assurance process and (ii) describe the erroneous values that they observed! This information will help PurpleAir users look for erroneous temperature and relative humidity data to exclude from their own analyses. However, I think it would be more appropriate for the information on the nature and frequency of the erroneous temperature and RH values to appear along with the information on how many data points were excluded due to disagreement between the channel A and B $\text{PM}_{2.5}$ concentrations (Section 3.1).

2. Lines 116–120: The results shown in Figure 1 are discussed here and in Section 3.1. I would prefer to see this discussion all in one place.

C6

3. Lines 121–127: If the “within 5 $\mu\text{g m}^{-3}$ or 61%” check was performed on the 2-minute/80-second averages, I think it would make more sense for this text to appear before the paragraph describing the completeness threshold. If the “within 5 $\mu\text{g m}^{-3}$ or 61%” check was performed on 24-hour averages, I think this text should remain here.
4. Lines 128–134: I think this text should appear before the paragraph on lines 102–109. The authors should describe how they downloaded the data before describing how they cleaned and averaged the data.
5. Lines 135–142: This material should appear after the first paragraph in Section 2.2.1 so that the authors’ description of how they compared channel A and B data for quality assurance will make sense to readers who are not already familiar with Plantower sensors or PurpleAirs.
6. Lines 219–220: I think this sentence should be moved to Section 2.2.1. Can the authors also provide additional information on this collocation? Where did it take place? How many PurpleAirs were collocated? I assume this was only monitors sent out by EPA and not monitors already being operated by air monitoring agencies. How long did the collocation last? What duration were data averaged over before calculating Pearson and Spearman correlations? What duration were data averaged over before comparing the channel A and B $\text{PM}_{2.5}$ concentrations?
7. Lines 242–251: I think it would be more appropriate for this paragraph to appear along with the material discussed in Section 2.2.
8. Line 255: I would prefer to see Table S4 in the main manuscript. In Table S4, why is the RH model just called “RH” but the temperature and dew point models are called “+T” and “+D,” respectively? Can the models be labeled RH, T, and D or +RH, +T, and +D? Also, why not group Nonlinear RH with RH, +T, and +D? Why not group +RH+T+D with +RH+T, +RH+D, and +D+T?
9. Lines 262–264: I think this sentence should appear with the description of Plantower

C7

data in Section 2.2.1.

10. If the authors want to stick with the conventional “methods” and “results” organization, the text on lines 255–256, 262–272, 276–277, 288–293, 299–300, 306, and 310–329 is technically methods. The models considered should be described in the methods section and then the amount of variance explained/comparison to prior studies should appear in the results section. If the authors wish to use some alternative organization, I think it would make sense to merge Section 2.3 with Section 3.3.

TECHNICAL CORRECTIONS

1. Line 12: Please insert a comma after “sensors” and another after “(PM)”.
2. Lines 45–50: This sentence contains five instances of “and” and zero commas, which made it difficult to follow. Please consider rephrasing or splitting it into multiple sentences.
3. Line 53: Insert “(Plantower PMS5003)” after “particle sensors”.
4. Line 54: Specify “(Bosch BME280)” instead of “(BME280)”.
5. Line 68: “location specific” should be “location-specific”.
6. Line 71 and elsewhere: “U.S. wide” should be “U.S.-wide”.
7. Line 98: Please insert a comma after “sensors” and another after “B”.
8. Lines 113–114: I suggest rewriting this sentence in past tense to be consistent with the rest of the paragraph (change “ensures” to “ensured” and “are” to “were”).
9. Lines 135–136: This sentence was difficult to follow. Maybe rephrase as: “The Plantower sensor reports estimated mass of particles with aerodynamic diameters $< 1 \mu\text{m}$ (PM_1), $< 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$), and $< 10 \mu\text{m}$ (PM_{10}).”
10. Lines 145–146: I think “earliest data” should be “date”.
11. Line 148: “local time” should be “local standard time,” correct?

C8

12. Lines 284–285: I assume “30 and 80%” refers to RH? Can you write “30% and 80% RH” to ensure this is clear?
13. Line 305: Insert a comma after “Lastly”.
14. Line 311: “corrections” should be “correction”.
15. Line 314: There are two periods at the end of this sentence.
16. Line 406: The authors alternate between “quality assurance,” “cleaning,” and “quality control” throughout the manuscript. More consistent terminology would be preferable.
17. Line 420–421: Should this sentence read, “In this work, we also excluded three PurpleAir sensors because there was overall poor agreement between the A and B channels even after excluding individual data points.”? If not, then maybe I do not understand what the authors are trying to say.
18. Line 562: Should there be an “and” between “correction methods” and “quality assurance methodology”?
19. Line 559–561: Can the authors please revise this sentence? It’s quite long and contains some redundant statements.
20. Figure 1: I like how the authors use this figure to illustrate possible data quality issues, but it’s really hard to see the data here. Can the authors make the whole figure larger, use smaller strip text so that the plot panels are larger, and use open circles as markers? Can the authors also either reorder the facets so that the PurpleAirs in CA, for example, appear in the order CA1, CA2, CA3, CA4, CA5, CA6, CA7, CA10, CA11, CA12, CA13, CA15, CA16, CA17, CA18, CA19? This could be accomplished by labeling “CA1” as “CA01” or by reordering the factor associated with these labels.
21. Table S7: There is a broken reference at the end of the caption.

C9

Interactive comment on *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2020-413, 2020.