The authors thank the reviewer for their second set of comments. We have modified the manuscript based on these comments to improve and clarify the text. Please find below our detailed responses in bold blue text (with direct quotes from the revised manuscript shown in "bold, quoted and italic" text). The reviewer's comments are shown in black unformatted text. All line numbers in our responses correspond to the "clean" version of the revised manuscript.

RESPONSE TO THE COMMENTS FROM THE REVIEWER

While some improvements have been made to this paper, in my opinion the authors have not yet adequately shown what their work adds to past work. They use a new correction method but do not compare to the accuracy that could have been achieved with other commonly used corrections in the literature. They only compare to the Barkjohn equation. The Barkjohn equation is one of the older PurpleAir corrections and multiple other corrections showing better performance than the Barkjohn equation have been published in the past 3 years (e.g., Wallace, Nilson). This was an issue I brought up on the last round of the draft that the authors did not address. I hope that the authors will address this comment along with my other specific comments below. They seem to have added additional inaccuracies in some places and may have errors in some figures. I do not feel this paper is ready for publication but think that it could be after another round of major revisions.

Major:
  1- On the last round of revisions, I and the other reviewer requested that the author consider other corrections common in the literature. To me considering other equations would mean applying corrections of similar form to their dataset and then comparing how the coefficients and performance compare to past work and other corrections. Instead, they have just added a paragraph summarizing the results from the past studies.

Response: We thank the reviewer for the comment. We edited the previous paragraph (lines 377-389) to apply other existing non-linear correction models to our PurpleAir sensors data and compare them with our developed models.

"
We compared our results with three nonlinear models that were previously tested for PurpleAir sensors. Two of these studies were not fit with data for our warm-humid climate zone study area. Malings et al. (2020) developed a two-piecewise linear model based on a threshold of 20 μg m^{-3} PM$_{2.5}$ concentrations using 11 PurpleAir sensors at 2 sites in Pittsburgh. The Malings et al. (2020) paper includes DP as one of the predictors (Table 3), which violates the assumption of predictor variable independence in the correction model since a high correlation was found between DP and T. Performance metrics for the Malings et al. (2020) model were inferior to those for our models and for the models developed by other authors (Table 3). Wallace et al. (2021, 2022) estimated correction factors based on the ratio of the mean AQS to the mean PurpleAir for all pairs of PurpleAir/AQS sites from California (Wallace et al., 2021), and from California, Washington and Oregon (Wallace et al., 2022) in separate models. Using the correction factor of 3 (ALT-CF3) recommended in Wallace et al.

(2021), we calculated higher MAE and RMSE (Table 3) than for any of our models and for the Barkjohn model. Similarly, the correction model developed by Nilson et al. (2022) to the cf=Atm data (same type of data used in their model) yielded similar R^2 and even higher RMSE and MAE than found with the ALT-CF3 model (Table S9). Nilson et al. (2022) used 35 PurpleAir/FEM sites in the U.S. and Canada including 2 sites in our study area.

Table 3: Other previously developed nonlinear correction models

Correction models		Model fit with hourly data			
		R^2 (%)	RMSE ($\mu g\ m^{-3}$)	MAE ($\mu g\ m^{-3}$)	R (%)
Wallace et al. (2021)	ALT-CF3	68	3.88	2.86	82
Nilson et al. (2022)	$pm25\_atm/(1 + 0.24/(100/RH - 1))$	68	4.14	2.98	82
Malings et al. (2020)	$75 + 0.60\ PA_i - 2.50\ T_i - 0.82\ RH_i + 2.9\ DP_i\ (for\ PA > 20\ \mu g\ m^{-3})$ $21 + 0.43\ PA_i - 0.58\ T_i - 0.22\ RH_i + 0.73\ DP_i\ (for\ PA \leq 20\ \mu g\ m^{-3})$	22	11.08	9.56	47

"

2- Met One BC-1060 – this is a black carbon monitor not a PM2.5 monitor. Why would you compare it to PM2.5? It looks like this site also runs an FRM R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC that would have been more appropriate to compare to (https://www.epa.gov/outdoor-air-quality-data/interactive-map-air-quality-monitors). I don't think these sites should be included in Figure 1. It is misleading when you didn't have any PM2.5 data you compared to in GA to include it in the states that you had comparisons for. At a minimum it should be in a different color if not included in the model development.

Response: Thank you for pointing that out. We double checked the monitor type. The monitor in GA mentioned in our study is indeed the R & P Model 2025 PM-2.5 Sequential Air Sampler w/VSCC. The type of monitor has been corrected in Table S13. We apologize for the error.

3- If you are using only hourly data, then you will never use FRM data since it is only 24-hr averages. You should clarify this throughout including "included only hourly average PurpleAir data points that had a spatial (within the calculated radius) correspondence to hourly FRM or FEM concentration". When you do the 24-hr analysis do you only use 1-hr data averaged up to 24-hr data or do you also pull in the 24-hr FRM data?

Response: Thank you for the comment. All of our AQS data, including the averaged 24-hr data, were initially hourly data. To address the comment, we edited the manuscript to add that detail (Footnote 1) and remove the "FRM" acronym in that statement (lines 150-152).

"Following data cleaning, the final PurpleAir concentration (C_{PA}) dataset used in our study was obtained by averaging Channels A and B and included only hourly average PurpleAir data points that had a spatial (within the calculated radius) correspondence to hourly FEM[1] concentration (C_{AQS}) data."

"Footnote 1: The AQS reference monitors used in our study were FEM monitors."

4- Figure 4: I am surprised that Model 4 performs better. It seems like the green plot shows more scatter and also has a lower slope further from the 1:1 line. I also don't see any negative data from the Barkjohn equation even though you stated that was an issue. Comparing Figure 4 to 7, I think one of these plots is wrong in Figure 7 model Bj is closer to the y axis while in Figure 4 Model 4 is closer to the Y axis. Am I missing something or is there a mistake in your figure labeling?

Response: We are very grateful to you for identifying this error. Upon further review, we found that the captions from Figure 4a and Figure 4b were interchanged. We edited the caption to correct the mistake (lines 332-334).

"Figure 4: Positive linear correlation between daily AQS and daily predicted PM₂.₅ concentrations with RH distribution (a) AQS and predicted PM₂.₅ concentrations using Model 4 of the MLR process shown in purple (b) AQS and predicted PM₂.₅ concentrations using the Barkjohn model shown in green."

Minor:
5- "However, we did not include Nilson et al. (2022) since they only developed linear models using CF-1 PurpleAir data." Their work was based on the cf_atm data based on the corrigendum they released last July https://amt.copernicus.org/articles/15/3315/2022/amt-15-3315-2022-corrigendum.pdf.

**Response: Nilson et al. (2022) has now been included. Please see our response to comment #1.**

6- Figure S4. It is hard to understand what this means since you haven't normalized for the true monitor concentration (i.e., dividing each hourly PurpleAir concentration by that hours monitor concentration)

Response: Thank you for the comment. Figure S4 has been edited to normalize the data.

"

---

[1] The AQS reference monitors used in our study were FEM monitors.
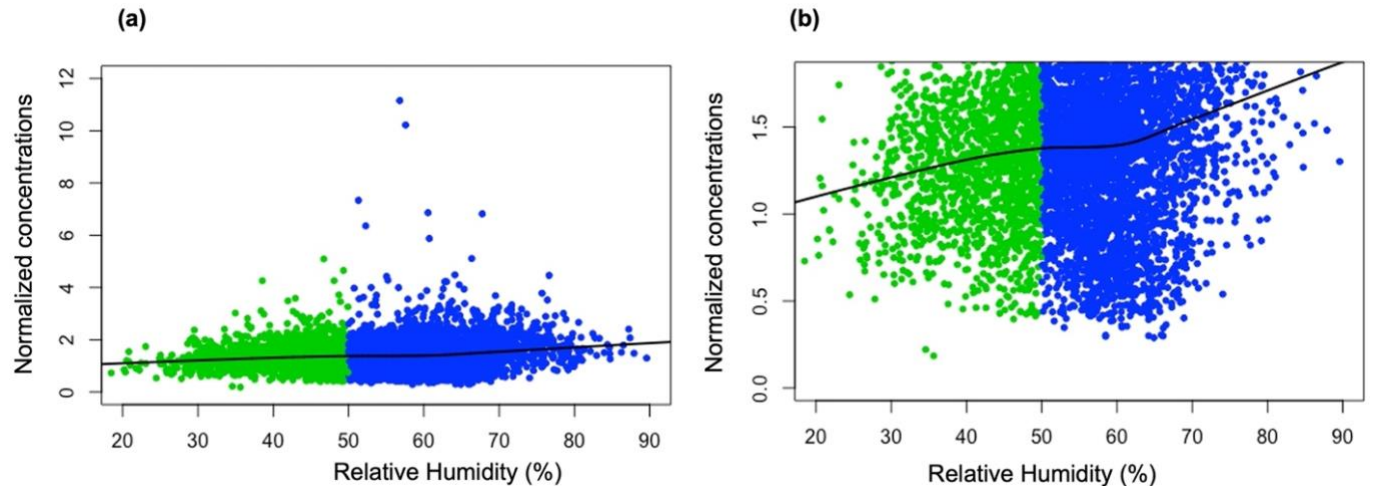
Figure S1: Correlation between the ratio of raw PM$_{2.5}$ PurpleAir and AQS concentrations and RH showing the nonlinearity of PM$_{2.5}$ PurpleAir concentrations. Graph a) represents the entire dataset, and graph b) is a zoom in to better display the regression line and the nonlinearity of the data."

7- The authors didn't check their dataset to see if there were any alternative PMS5003s they just said that could have impacted their results when they easily could have checked the ratio between the bin data to see if they had any of the alternative PMS5003s.

Response: Thank you for the comment. We edited the previous paragraph to specify how many alternative PMS5003s were found among our sensors (lines 294-299).

"Additionally, a study conducted by Searle et al. (2023) found that 12.9 % of the sensors deployed by PurpleAir between June 2021 and May 2023 reported negative bias of approximatively 3 µg m-3. These PurpleAir sensors, specifically deployed between June 2021 and January 2022 and between March to May 2023, used an alternative Plantower PMS5003, which affected the reported particle size distributions and concentrations (Searle et al., 2023). Based on the technique developed by Searle et al. (2023) to identify PMS5003 sensors, we estimated that only one of our sensors (sensor ID: 116559), representing 0.62% of our data, fell into this category. This may have a slight effect on the performance of our models."

8- "The LOSOCV resulted in a RMSE and a MAE of 3.31 µg m-3 and 2.29 µg m-3 respectively for Model 4. These values were higher than those for the LOGOCV process, which is not surprising considering the variability between states." 3.32 is the value in table S8.

Response: Thank you for the comment. The typo has been corrected in the manuscript (lines 303-304).

"The LOSOCV resulted in a RMSE and a MAE of 3.32 μg m⁻³ and 2.29 μg m⁻³ respectively for
Model 4."

9- "To ensure data accuracy, AQS data are collected by FRM or FEM, which are typically
filter-based monitors (U.S. EPA, 2023b)" I'm not sure this is helpful. While FRM are
filter based and a beta attenuation monitor uses filter tape, the T640 is optical and is the
most common FEM used in your study.

**Response: Thank you for the comment. We edited the manuscript to remove the filter-
based statement (line 127).**

"To ensure data accuracy, AQS data are collected by FRM or FEM (U.S. EPA, 2023b)."

10- "Based on the electronic effects of water uptake" Is this accurate? I thought this was due
to the particles up taking water.

**Response: Thank you for the comment. We edited the manuscript and restated the
sentence (lines 164-166).**

"Because measurement errors are related to water uptake by particles (Hagan & Kroll, 2020;
Rueda et al., 2023; Wallace et al., 2021), temperature (T) and relative humidity (RH) are the
most commonly found bias correction parameters in the literature (Ardon-Dryer et al., 2020;
Bi et al., 2020; Magi et al., 2020; Malings et al., 2020; Wallace et al., 2021) for the
PurpleAir."

11- Table S13 – it would be helpful to include the number of valid data points for each pair to
better understand what part of the time period is represented since there is so much
missing data according to Tables S1.

**Response: Thank you for the comment. The number of valid data points for each pair of
PurpleAir/AQS has been added to Table S13. Figure 2, produced after the QA process,
also shows what time period is represented by each pair of data.**

12- Line 383 (track changes version): as I stated in my first review Zheng et al. is not about
PurpleAir sensors PurpleAir should be replaced with Plantower in this sentence.

**Response: We appreciate the comment. We edited the sentence to replace "PurpleAir" with
"Plantower." (lines 308-311)**

"Zheng et al. (2018), evaluating the performance of Plantower PMS3003, which is similar to
the PM₂.₅ sensor used in PurpleAir, found an R² value of 66 % for a 1-h averaging period after
applying an MLR calibration equation to compare three Plantower sensors against each other
and a co-located reference monitor over a period of 30 days."

13- Line 421: SD is not listed in Table 1. Is it an error metric?

Response: We appreciate the comment. The sentence was restated to remove the confusion (lines 329-330).

"An evaluation of Model Bj applied to our warm-humid climate zone daily PurpleAir datasets revealed substantially higher error metrics than the other models (Table 1)."

14- Line 489: "We compared our results with some nonlinear models that were previously tested for PurpleAir sensors" I don't think you have done that. You haven't even drawn the conclusion for the reader on whether their error is higher or lower than yours and you didn't apply them to your dataset.

Response: Thank you for the comment. Please see our response to comment #1.

15- Line 465: " Click or tap here to enter text." Remove

Response: Thank you for the comment. This erroneous text has been removed.