The authors would like to thank the editor, the two reviewers, and Dr. Ouimette for their thoughtful and thorough review, and constructive remarks. We have modified the manuscript based on these comments to improve and clarify the text. Please find below detailed responses in bold blue text (with direct quotes from the revised manuscript shown in "bold, quoted and italic" text) to the comments and suggestions offered by the reviewers (shown in normal text). All line numbers in our responses correspond to the "clean" version of the revised manuscript.

RESPONSE TO THE COMMENTS FROM REFEREE 2

General comments:

This paper provided the evaluation of PurpleAir correction using the warm, humid climate zones data and aimed to improve the EEPA Barkjohn model. It provides helpful information about improved performance metrics and avoids collinearity using DP, RH and T. However, the multilinear regression has been used before. There is no significant scientific insight gained with the new parameters. Several suggestions to strengthen this paper:

Response to the general comments: The authors appreciate the reviewer general comment on the scientific insight. However, we respectfully disagree with the comment. The objective of the paper was to develop and evaluate PurpleAir bias correction models (a more accurate model) for use in areas under high humidity conditions considering the sensitivity of PurpleAir sensors to humidity. Moreover, our study evaluated the performance of MLR models and a novel semi-supervised clustering method as a model-based clusters (MBC).

"The objective of this study is to develop and evaluate PurpleAir bias correction models for use in the warm humid climate zones (2A and 3A) of the U.S. (Antonopoulos et al., 2022). First, we tested an MLR with different combinations of predictive variables. To avoid the transferability constraints observed for the GMR, our study then tested a novel semi-supervised clustering method. We used PurpleAir data and the FRM/FEM $PM_{2.5}$ data from the EPA Air Quality System (AQS) database from January 2021 to August 2023. We tested new correction models developed for the high-humidity Southeastern region of the country and compared them with the EPA nationwide PurpleAir data correction model proposed by Barkjohn et al. (2021)." **(lines 75-80)**

1- Consider other correction methods and explain what can provide the best insight of the Purpleair data.

Response: We thank the reviewer for this comment. In addition to the two methods tested in our study and their comparison with the model developed by Barkjohn et al. (2021), a paragraph was added to the manuscript to compare the results of our study with other non-linear models previously used (lines 369-377). Please see our response to Referee 1, comment #1 copied below.

Response to Reviewer #1: The authors appreciate the reviewer's suggestion. We added a new paragraph (lines 369-377 in the Results and Discussion section) to compare the models developed in this study with other existing non-linear models as suggested. However, these models were designed for specific locations and not intended to work for a broad area. Moreover, none of these studies covered the Southeastern U.S. Malings et al. (2020) used data from 2 sites in Pittsburgh. Wallace et al. (2022) used data from California, Washington and Oregon. We added the results found by Wallace et al. (2001, 2022) and Malings et al. (2020). However, we did not include Nilson et al. (2022) since they only developed linear models using CF-1 PurpleAir data.

"We compared our results with nonlinear models that were previously developed and tested for PurpleAir sensor bias correction. Malings et al. (2020) developed a two-piece linear model based on a threshold of 20 µg m-3 PM2.5 concentrations using 11 PurpleAir sensors in 2 sites in Pittsburgh. The models included CPA, T, RH and DP as predictors. They found a correlation below 50 % and a MAE ranging from 3 to 5 µg m-3 (Malings et al., 2020). Some other studies (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 first using 33 PurpleAir sensors from California (Wallace et al., 2021) and then including 182 PurpleAir sensors from California, Washington and Oregon (Wallace et al., 2022). Their studies evaluated alternative PM2.5 PurpleAir estimates, however Wallace et al. (2021) also developed a correction factor for the cf_1 PM2.5 PurpleAir estimates. They calculated a range of a correction factors between 0.65 and 0.72 resulting in an overestimation of PM2.5 of 40 % compared with AQS monitors (Wallace et al., 2021)." **(lines 369-377)**

2- Typically, the low-cost sensors measure the PM base on the optical size, and it is unclear how they can accurately predict the aerodynamic size and get the correct PM2.5. The conversion of particle aerodynamic size to optical size, or vice versa, is not straightforward because it depends on several factors, including the particle's shape, density, and refractive index. Are the FRM/FEM monitors filter-based measurements? How does the linear regression provide reliable information?

Response: We thank the reviewer for the comment. FRM/FEM monitors are reference-grade monitors designated by EPA. EPA has evaluated every FRM/FEM to ensure that it is producing accurate concentrations based specific standards (40 CFR Appendix L to Part 50). Moreover, we have already pointed out in the manuscript that optical sensors have many challenges in accurate detection of PM2.5 (lines 47-50).

"Most low-cost PM sensors, including the PurpleAir sensor, utilize optical sensors based on the light-scattering principle to estimate PM mass concentration. Thus, they are subject to measurement errors from various factors, including particle size, composition, optical properties, and interactions of particles with atmospheric water vapor (Hagan & Kroll, 2020; Rueda et al., 2023; Zheng et al., 2018; Zusman et al., 2020)." **(lines 47-50)**

With regard to the reliability of the modeling method, the linear regression is designed to correct less accurate PA sensors based on the more accurate AQS monitors. The performance of a linear regression is measured in general by its precision of linearity using R^2 and R and by the accuracy of the error metrics. The performance metrics evaluated in our study are presented in lines 210-217.

Specific comments:

3- Line 127-129, Please explain how to determine the detection range for PurpleAir? The reference used 1.15-2.55? This paper used 1.5? Why not 1.6? or 1.75?

Response: We appreciate the comment. One of the references was missing. We clarified the statement and added the missing reference in lines 134-136.

"We applied a series of data exclusion criteria for quality control. First, we used a detection limit of 1.5 $\mu g\ m^{-3}$ for the PurpleAir data. This value is equivalent to the average of the values reported by Tryner et al. (2020) and Wallace et al. (2021) for the cf_1 data series." **(lines 134-136)**

4- Line 131, What is the difference between the two channels? Should we expect them to agree in a certain percentage at low and high concentrations?

Response: We thank the reviewer for expressing the concern. There is no difference between the design of the 2 channels. They are both PM2.5 sensors arbitrarily designated as Channels A and B. We edited the sentence to add the word *"arbitrarily"* for more clarity (line 138).

The data cleaning criteria for the agreement between the 2 channels for both low and high concentrations are already defined in the manuscript in lines 136-146.

5- Line 141, For each site, how much data remained? Does this data cleaning cause any bias in the data collection?

Response: We thank the reviewer for the comment. Fig. S1 presents the number of data points remaining to be used in the study per site (n from Fig. S1 corresponds to the number of data points per PurpleAir site). Moreover, we added a table (Table S1) to present how much data were removed in the process at each step.

The role of the data cleaning is to minimize biases in the modeling process.

"The QA process removed about 22 % (Table S1) of the raw data…" **(line 236)**

"Table S1: Percentage of hourly data removed by QA process from the initial 56 PurpleAir sensors

QA criteria	*% removed**
Process 1: Removing NAs (PM, T, RH)	*2.026*
Process 2: Channels A & B agreement	
Low concentration (≤ 25 μg/m³): 537,246 obs.	*2.242*
High concentration (>25 μg/m³): 80,196 obs.	*2.056*
Process 3: A & B concentration < 1.5 μg/m³	*6.753*
Process 4: Average A & B concentration > 1000 μg/m³	*0.005*
Process 5: Removing data from sensors with RH issues	*5.527*
Process 6: Removing RH≠ 0-100% and T ≠ 0-130 ºF	*3.484*

percent removed from the total number of observations"

6- Section 2.4.2, the equations are confusing. Will the beta 2 in equation 2 be the same as the beta 2 in equation 3?

Response: We thank the reviewer for expressing the concern. The equations follow the general mathematical notation of a multilinear regression model (see equation below). Each beta is the regression coefficient of a predictor X, whose name is defined in the equation (C_{PA}: PurpleAir $PM_{2.5}$ concentration, RH: relative humidity, T: temperature). They will not have the same values.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

7- Table 1, the parameters from each model have a very high precision. Is it realistic to include such high precision?

Response: We thank the reviewer for the comment. We included such a high precision with many significant figures so that users of our models would not have rounding errors in their datasets.

8- Figure 4 the data plotted seemed to be from two groups. One follows 1:1 line, and the other one follows 2:1 line. Cluster 2 still has the 2:1 group. Is there any other reason for this 2:1 group?

Response: We thank the reviewer for the interesting observation. We acknowledge that this represents an area of uncertainty. We added a sentence in the Results and Discussion section (lines 322-324 and lines 383-385) to acknowledge the cluster formation as a limitation.

***"An aggregate of datapoints can be seen on the left-hand side of the correlation plots (Fig. 4) to deviate from the model fit line. These data probably influenced the performance metrics of the models."* (lines 322-324)**

"The same aggregate of datapoints seen in Fig. 4 is also observed in the SSC models, but only in Cluster 2 (Fig. 6). This may have affected the accuracy of the model (Table 1)." **(lines 383-385)**