We thank the reviewer for their time reviewing our manuscript and their feedback. We have taken the time to make a number of significant improvements to the manuscript based on the reviewers' comments. We have included responses in blue below.

Anonymous Referee #3

Received and published: 9 December 2020

The manuscript presents a development of a generalized correction equation for the PurpleAir $PM_{2.5}$ sensors in the USA. The used data set includes 39 different sites in 16 states and 50 sensor units altogether. The number of recorded datapoints was 12,635 (24h-average, from 8 to 3762 per state). Only sites which had reference level data available no further than 50m away were included in the analysis.

Major comment:

The chosen analysis approach does not utilize the data as well as it could, and there is a strong case to be made that the generalized correction equation is not representative: in fact, a single generalized and true correction equation is most probably impossible to form due to the site-specific differences in aerosol composition. Is it not obvious that two different sites entailing two completely different environments (e.g. rural background vs. urban city center) are not representative of each other? To add, the analysis does not consider possible seasonality, which may have a significant impact on aerosol composition: the shortest sensor data is composed of only 8 data points (MT, Table 1). When these issues are coupled with the notion that not all the factors affecting the sensor behavior are fully understood (line 296), generalizations should be made with extreme caution. In my opinion, the only way to address this issue is to generate correction equations for a few, most generalizable environments (e.g. urban, urban background, detached housing area, regional background) with known aerosol sources. This would also lend for a further investigation regarding the underlying reasons affecting sensor responses.

While we acknowledge that future work could dig more into further reducing error and specific projects could make more accurate corrections for small areas that is not the objective of this project. While site- or location-specific corrections may have merit, this dataset, which to our knowledge is the most comprehensive published in the literature to date, is insufficient to generate these types of localized corrections for use nationally. We believe our dataset is comprehensive enough, including the range of meteorological conditions and aerosol compositions, to build a generalized correction for the entire U.S. that is fairly accurate and able to be implemented easily. We have shown in this paper that this generalized correction is similar to past corrections in the literature suggesting that it is representative of previously studied areas as well. The developed correction equation is valuable for U.S.-wide correction of PurpleAir sensors. If we tried to apply a correction based on environment or seasonality across the U.S. it would be impossible since this information is unknown and there are too many gaps in our dataset.

Technical comment:

In multiple linear regression independent variables should be independent. In my opinion, the fact that previous studies have misused linear regression does not warrant for a new research to continue misusing it. It reinforces the bad habit and undermines the quality and significance of the whole research line of low-cost aerosol sensors. Besides the relative humidity and temperature analysis, this applies also for the analysis regarding the binned particle counts (line 270).

> We have removed all equations that have strongly correlated additive terms in the discussion. See the updated set of equations considered in Table 2:

Table 2. Correction equation forms considered, and the root mean squared error (RMSE). The best performing model from each increasing complexity (as indicated with *) was validated using withholding in the next sections (Sections 4.3.2 and 4.4).

Name	Eqn	RMSE ($\mu g\ m^{-3}$) (cf_1)	RMSE ($\mu g\ m^{-3}$) (cf_atm)
linear	$PA=PM_{2.5}*s_1+b$	2.88*	3.01
+RH	$PA = s_1*PM_{2.5} + s_2*RH + i$	2.52*	2.59
+T	$PA = s_1*PM_{2.5} + s_2*T + i$	2.84	2.96
+D	$PA = s_1*PM_{2.5} + s_2*D + i$	2.86	2.99
+RH*T	$PA = s_1*PM_{2.5} + s_2*RH + s_3*T + s_4*RH*T + i$	2.52	2.60
+RH*D	$PA = s_1*PM_{2.5} + s_2*RH + s_3*D + s_4*RH*D + i$	2.52	2.60
+D*T	$PA = s_1*PM_{2.5} + s_2*D + s_3*T + s_4*D*T + i$	2.51*	2.61
+RH*T*D	$PA = s_1*PM_{2.5} + s_2*RH + s_3*T + s_4*D + s_5*RH*T + s_6*RH*D + s_7*T*D + s_8*RH*T*D + i$	2.48*	2.57
PM*RH	$PA = s_1*PM_{2.5} + s_2*RH + s_3*RH*PM_{2.5} + i$	2.48*	2.53
PM*T	$PA = s_1*PM_{2.5} + s_2*T + s_3*T*PM_{2.5} + i$	2.84	2.96
PM*D	$PA = s_1*PM_{2.5} + s_2*D + s_3*D*PM_{2.5} + i$	2.86	3.00
PM* Nonlinear RH	$PA = s_1*PM_{2.5} + s_2\frac{RH^2}{(1-RH)}*PM_{2.5} + s_3*\frac{RH^2}{(1-RH)} + i$	2.86	2.99
PM*RH*T	$PA = s_1*PM_{2.5} + s_2*RH + s_3*T + s_4*PM_{2.5}*RH + s_5*PM_{2.5}*T + s_6*RH*T + s_7*PM_{2.5}*RH*T + i$	2.46*	2.53
PM*RH*D	$PA = s_1*PM_{2.5} + s_2*RH + s_3*D + s_4*PM_{2.5}*RH + s_5*PM_{2.5}*D + s_6*RH*D + s_7*PM_{2.5}*RH*D + i$	2.54	2.57
PM*T*D	$PA = s_1*PM_{2.5} + s_2*T + s_3*D + s_4*PM_{2.5}*T + s_5*PM_{2.5}*D + s_6*T*D + s_7*PM_{2.5}*T*D + i$	2.52	2.63
PM*RH*T*D	$PA = s_1*PM_{2.5} + s_2*RH + s_3*T + s_4*D + s_5*PM_{2.5}*RH + s_6*PM_{2.5}*T + s_7*T*RH + s_8*PM_{2.5}*D + s_9*D*RH + s_{10}*D*T + s_{11}*PM_{2.5}*RH*T + s_{12}*PM_{2.5}*RH*D + s_{13}*PM_{2.5}*D*T + s_{14}*D*RH*T + s_{15}*PM_{2.5}*RH*T*D\ i$	2.42*	2.51

We have also added correlograms and discussion of them to the text (see response to reviewer 2 above).

Another misuse particularly characteristic for low-cost sensor studies is the use of R2 as goodness-of-fit indicator in nonlinear regression. R2 is not valid for nonlinear regression (line 295).

We have updated the text to consider RMSE instead of R^2 (Table 2).

Recommendation: reject

It is our hope that the extensive edits and improvements to the updated manuscript in response to the numerous thoughtful comments from all reviewers will change this recommendation.