Response to Reviewer #2

Kudos to the author team for putting together this impressive paper estimating daily surface-level PM2.5 in Ghana combining a number of satellite- and model-derived datasets and an impressively large in the African context and multi-sensor/diverse network of low-cost instruments through machine learning.

The writing is excellent; figures are clear (I especially like figure 1 and think it does a good job at distilling the complicated procedure for generating the authors' dataset); and I appreciate the US-Ghanaian partnerships demonstrated through this paper's slate of authors and the deployment of the in-situ monitoring network leveraged within. To me, this study is well on track for publication in AMT. I have included several comments below that, when addressed, should make the paper in excellent shape for publication.

Introduction:

- Lines 35ff: The author might consider, in addition to mentioning the overall burden of disease from air pollution in Africa, stating the estimated burden in Ghana alone. I believe these values should be readily accessible via the Global burden of Disease's/IHME's Global Health Data Exchange (GHDx). It might help contextualize the importance of this particular word better.

  Author response: We thank the reviewer for this helpful suggestion. We have added a Ghana-specific estimate of the health burden attributable to air pollution in the *Introduction* section to provide additional context for the importance of the study. The estimate is now included in the Introduction immediately before introducing Ghana as the study domain.

  Revision: We added the following sentence to the Introduction and cited the Global Burden of Disease estimates published by IHME for 2023.

  "*In Ghana, exposure to air pollution was estimated to contribute to around 32,500 deaths in 2023, with nearly all of these deaths (32,400) attributed to particulate matter pollution, underscoring the substantial health burden associated with air pollution in the country.*"

- Lines 85ff: The sentence beginning with "Furthermore, …" feels a little misplaced. Prior to this, the authors are discussing recent advances in high resolution PM2.5 modeling, so it feels weird to pivot back to the deficiencies of MERRA-like products. This sentence might be better suited earlier in the Introduction when discussing challenges of limited monitoring.

  Author's response: We agree that the discussion of the spatial limitations of existing MERRA-2-based $PM_{2.5}$ products is better placed alongside the earlier discussion of limitations in available air-quality observations, especially in satellite-based products. We have therefore moved this statement to the earlier section of the Introduction to improve the logical flow.

Revision: We have moved the sentence to the end of the earlier paragraph discussing satellite-based air-quality products and revised it as follows:

"*Furthermore, although existing global $PM_{2.5}$ products derived using Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2) reanalysis parameters provide continuous spatiotemporal coverage, the spatial resolution remains too coarse (>50 km) to capture intra-city $PM_{2.5}$ variations across most cities in Africa.*"

Methods:

- Section 2.2, Figure S2: The low-cost sensors have uneven spatiotemporal coverage and are available only from 2018 onwards. Are there ways that the co-variations in AOD and other measures of pollution, meteorology, and other inputs to their machine learning model are unique to the 2018-2025 period (b/c climate change, b/c teleconnections, etc.) such that applying these machine-learned insights to previous periods spanning 2005-2018 could be inappropriate if the coupling among these drivers have changed, especially in nonlinear ways? It would be good for the authors to discuss this potential and even better if there are quantitative ways to demonstrate the impact that this could have on results (assuming I'm thinking through it correctly).

  Author's response: We addressed the concern regarding the potential for temporal changes in predictor–$PM_{2.5}$ relationships by evaluating temporal generalizability using leave-one-year-out cross-validation across the available 2018–2024 observations. The model maintained consistent predictive performance across the held-out years, providing evidence of temporal robustness. We have also clarified that the estimates for 2005–May 2018 cannot be directly validated against surface observations because in situ measurements were only available from August 2018 onward. These clarifications and validation results have been added to Sections 2.2, 2.5, and 3.2.

  Revision:

  <u>Section 2.2</u>

  "*$PM_{2.5}$ measurements collected from this sensor network between August 2018 and October 2024 were used as ground truth for model development. Since in situ measurements are available only from August 2018, $PM_{2.5}$ estimates for 2005–May 2018 were generated using models trained on the available surface observations and therefore cannot be directly validated against surface measurements from that period.*"

Section 2.5

"*This approach assesses the ability of the model to generalize across interannual variability in meteorology, emissions, and episodic pollution events, increasing confidence in the temporal robustness of $PM_{2.5}$ estimates for periods outside the training data. This step is particularly important because model training is based on observations from 2018–2024, while $PM_{2.5}$ concentrations are estimated retrospectively back to 2005, extending the application of the model beyond the period represented by the ground observations*."

Section 3.2

"*Under temporal cross-validation, XGBoost achieved R^2 values of 0.57, 0.56, and 0.59 for the TROPOMI, OMI, and MERRA models, with RMSE of 15.6, 15.6, and 15.2 μg m^{-3}, respectively. The leave-one-year-out validation results show that the model maintained consistent predictive performance across different years, indicating that the learned predictor–$PM_{2.5}$ relationships are not strongly dependent on the meteorological and pollution characteristics of any single year.* "

- In Sections 2.3.1, 2.3.2, and 2.3.3, I appreciated the detailed information about how the authors applied the QA/QC flags recommended for the different products to their study. It would be beneficial for readers to understand the fractions of data filtered out by these flags to better understand how often the ML methods rely on the MERRA model. Perhaps a schematic akin to Figure S2 could illustrate this nicely (although I understand it's a little more difficult than replicating Fig. S2 exactly since there's a spatial dimension to the satellite measurements). Relatedly, the authors allude to missing satellite data in Lines 369ff but I don't see any quantitative information on how common this is.

  Author's response: We acknowledge that quantifying satellite data availability provides useful context for understanding the contribution of the MERRA-2 fallback model. We evaluated the availability of satellite predictors across the 1-km grid after application of the product-specific QA/QC procedures. During the OMI period (2005–May 6, 2018), satellite predictors were available for an average of 58.3% of grid cells, with 41.7% requiring the MERRA-2 fallback. Following the introduction of TROPOMI on May 7, 2018, satellite availability increased substantially, with 99.1% of grid cells available on average and only 0.9% requiring MERRA-2 fallback. We also found a seasonal pattern in satellite availability, with greater reliance on the MERRA-2 fallback during periods of lower satellite coverage. These results are now provided as an annual summary in Table S11 and as a monthly heatmap in Figure S10.

  Revision: We added the annual satellite availability and MERRA-2 fallback fractions to Table S11 and the monthly spatial pattern of satellite availability to Figure S10. We also added the corresponding quantitative results to the Results section.

- Can the authors comment on the appropriateness of generating their PM2.5 product at ~1km2 when all the native inputs besides, I believe, AOD appear to be in the 5km2 or greater range? Unlike with some LUR products I've seen that have several fine scale (< 1km2) inputs like road locations and NDVI, this product seems to inherent all of the fine-scale features from AOD alone, which might be decoupled from surface-level PM.

  Author's response: The 1-km grid was selected to retain the fine-scale spatial variability represented by the MODIS MAIAC AOD while integrating complementary information from coarser-resolution atmospheric composition and meteorological predictors. Fine-scale predictors such as road networks, land-use characteristics, and NDVI could potentially provide additional local-scale information, as used in some land-use regression approaches in previous studies; however, the availability of consistently mapped fine-scale datasets remains limited, particularly across African countries, when considering the large spatial extent of a national-scale study.

  Our 1-km estimates integrate information across multiple spatial scales rather than assuming that all predictors independently resolve variability at 1 km. We have clarified this point in the revised manuscript.

  Revision: The following text was added at the end of Section 2.4.1.

  "*The 1-km prediction grid allows the model to retain the fine-scale spatial variability represented by the MAIAC AOD, such that adjacent grid cells can have different $PM_{2.5}$ estimates, while the coarser-resolution predictors provide complementary information on regional atmospheric composition and meteorological conditions. Thus, the resulting 1-km $PM_{2.5}$ estimates integrate information across multiple spatial scales, with MAIAC AOD providing fine-scale spatial information and the other predictors characterizing broader-scale atmospheric conditions.*"

- Line 387: I believe the authors meant to reference Figure S6, not S4, here.

  Author response: We have corrected the figure reference to Fig. S6.

  Revision: The revised sentence reads:

  "*The distribution of observed daily $PM_{2.5}$ concentrations is strongly right skewed, characterized by a long tail toward higher concentration values, as shown in Fig. S6 of the Supplementary Materials.*"

- Section 2.6: The use of 1-9km buffers to define the extent of cities feels overly simplistic. I am not familiar with most of the cities studied here, but I assume some might be bigger than 18km from north to south or east to west? I also just looked up Kintampo on Google Earth for the fun of it, and its built-up area appears to be quite elongated and not a circle. The authors might consider using products like GHS-UCDB or GHS-SMOD to more precisely define areas over which they will average PM2.5. Alternatively, if any Ghanaian agencies have city shapefiles/boundaries, those would be great too.

  Author response: We acknowledge that fixed circular buffers may not adequately represent urban areas with different sizes and spatial morphologies, particularly for elongated urban areas such as Kintampo. We therefore revised the approach used to define the spatial extent for $PM_{2.5}$ averaging. The selected 1-km grid cells are now identified using GHS-UCDB urban-center boundaries, where available, together with assessment of contiguous built-up development and the surrounding spatial context. This allows the spatial extent of the $PM_{2.5}$ averaging area to vary according to the size and morphology of each study location.

  Revision: The city-level $PM_{2.5}$ aggregation procedure in Section 2.6 has been revised accordingly, and the detailed spatial selection procedure is provided in Text S2. The GHS-UCDB dataset has also been added as a reference.

  <u>Main text (Section 2.6)</u>

  "*Daily $PM_{2.5}$ concentrations were derived for each of the 20 study locations by spatially aggregating the 1-km gridded $PM_{2.5}$ estimates over a location-specific set of grid cells selected according to the spatial extent and morphology of the surrounding built-up area (Text S2; Fig. S7). Where available, the Global Human Settlement Layer Urban Centre Database (GHS-UCDB)[41] urban-center boundaries were used as the primary spatial reference, with adjacent grid cells evaluated based on continuity of built-up development and surrounding spatial context. This approach accounts for differences in the spatial extent and morphology of the areas surrounding the 20 study locations. The selected cells were then used to calculate a representative daily $PM_{2.5}$ time series for each location.*"

  <u>Supplementary Information (Text S2)</u>

  "***Text S2. Selection of $PM_{2.5}$ grid cells for urban locations***

  To obtain a representative $PM_{2.5}$ time series for each of the 20 study locations, 1-km $PM_{2.5}$ grid cells representing each location were identified based on the spatial extent and morphology of the built-up area. Where available, the Global Human Settlement Layer Urban Centre Database (GHS-UCDB)[4] was used as the primary spatial reference for characterizing the spatial extent of each urban center. The GHS-UCDB urban-center boundaries, 1-km $PM_{2.5}$ grid, and city reference locations were overlaid to assess the spatial correspondence between the urban centers and the $PM_{2.5}$ grid.

For locations with a corresponding GHS-UCDB urban center, grid cells intersecting the GHS-UCDB urban-center boundary were used as the core set of cells. Adjacent grid cells were subsequently evaluated based on the continuity of built-up development and the surrounding spatial context to identify cells representing the broader area associated with each study location. Cells were included when they represented a contiguous extension of the built-up area associated with the study location, whereas cells separated from the developed area by clearly undeveloped or predominantly rural land, or containing isolated development, were excluded (Fig. S7). For locations without a dedicated GHS-UCDB urban-center boundary (Accra, Damongo, Goaso, Nalerigu, Navrongo, Sefwi Wiawso, and Tema), the city reference location and surrounding built-up morphology were used to identify representative grid cells using the same criteria.

This approach allowed the spatial extent of the selected $PM_{2.5}$ grid cells to vary according to the size and morphology of the built-up area surrounding each study location. The final selected grid cells were used to calculate the representative $PM_{2.5}$ time series for each location.

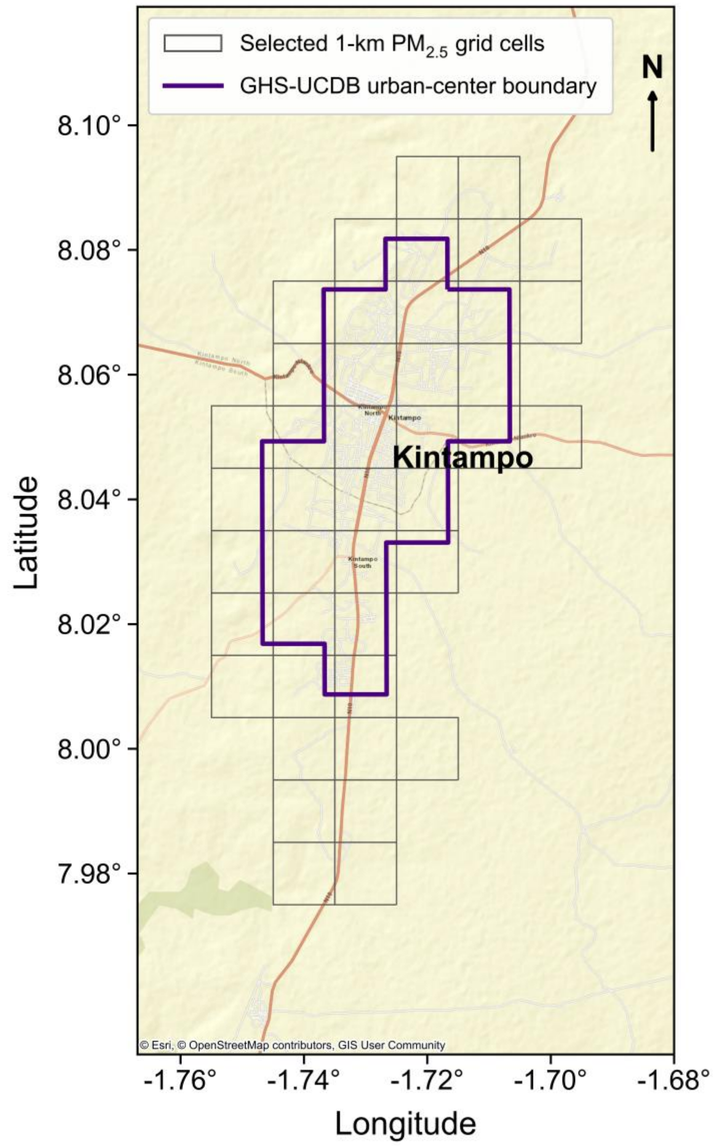

Fig. S1. Spatial selection of PM2.5 grid cells for Kintampo. *The 1-km PM2.5 grids are indicated by the gray grid lines and the GHS-UCDB urban-center boundary is shown by the thick purple outline.*"

- Sections 2.7.1, 2.7.2, and 2.7.3: I appreciate the detailed information about the statistical procedures but, in my opinion, feel like it could be shortened significantly or even removed. These statistical tests are fairly standard, and unless the authors modified them in some major way, I don't know readers need almost 2 pages devoted to them.

  Author's Response: We have substantially condensed Section 2.7. The revised section now briefly describes the application of the Seasonal Mann–Kendall test and Sen's slope estimator to the monthly $PM_{2.5}$ time series, the rationale for accounting for the strong seasonal cycle, the Benjamini–Hochberg false discovery rate correction for pixel-wise analyses, and the national and sub-regional trend analyses. Additional methodological details and equations are moved to Text S4 of the Supporting Information to ensure reproducibility.

Revision: Section 2.7, *Long-term $PM_{2.5}$ trend analysis*, has been substantially shortened and revised to address the comment. Detailed statistical formulations have been moved to Text S4 of the Supporting Information.

"*To quantify the direction, magnitude, and statistical significance of long-term changes in surface $PM_{2.5}$ across Ghana, we applied the Seasonal Mann–Kendall (SMK) test and Sen's slope estimator to monthly mean $PM_{2.5}$ concentrations derived from the multi-model output for 2005–2025. Monthly means were calculated from daily 1-km gridded $PM_{2.5}$ concentrations, yielding 252 monthly observations per grid cell. The SMK test was used to account for the strong recurring seasonal cycle in $PM_{2.5}$ concentrations, particularly during the Harmattan season.[42] Sen's slope estimator was used to quantify the magnitude of the monotonic trend in $\mu g\ m^{-3}\ yr^{-1}$.[43]*

Because the pixel-wise trend analysis involves spatially contiguous grid cells, we also evaluated spatial autocorrelation in the gridded $PM_{2.5}$ fields using Global Moran's I with a first-order rook neighborhood.[45] Global Moran's I was calculated annually for the 0.01° × 0.01° $PM_{2.5}$ fields for 2005–2025 and compared with the gridded $PM_{2.5}$ product of Shen et al. (2024) at 0.01° and 0.1° resolutions. Additional details are presented in Text S3.

For the pixel-wise trend analysis, the Benjamini–Hochberg false discovery rate (FDR) procedure was applied to the SMK p-values across all valid grid cells, with an FDR threshold of 0.05 used to identify statistically significant trends (Benjamini and Hochberg, 1995).[44] The SMK test and Sen's slope were applied independently to the monthly mean $PM_{2.5}$ time series at each 1 km grid cell across Ghana to produce spatially resolved maps of trend magnitude and significance. Additionally, spatially averaged time series were computed for the national domain and for distinct sub-regions of Ghana (northern, middle, and southern belts) to characterize large-scale changes in $PM_{2.5}$ burden over the 21-year study period. Further details of the pixel-wise trend analysis and FDR correction are provided in Text S4 of the Supporting Information."

Results:

- For Figure 2 and the associated discussion, while the MB improves nearly 20-fold for the TROPOMI model compared with the MERRA model, the MAE, RMSE, and correlation are nearly identical and not that much different from the OMI model. I think the authors attributed this similar performance to the complete daytime estimates from MERRA compared with the early afternoon measurements from TROPOMI and OMI, but could it also be that that PM2.5 is mostly explained by regional phenomena that are captured by the coarse MERRA product rather than needing the finer scale data provided by TROPOMI? Or does it potentially relate back to my earlier comment about the lack of fine-scale inputs?

Author's response: We agree that the comparable overall performance of the three model configurations suggests that a substantial component of daily $PM_{2.5}$ variability in Ghana is governed by regional atmospheric processes that can be captured by reanalysis-based predictors. This interpretation is also consistent with the SHAP analysis, which indicates that meteorological and seasonal predictors are among the dominant contributors to model predictions. However, the comparable performance in Figure 2 is derived from random cross validation and does not imply that the three model configurations capture equivalent spatial information. Under spatial cross-validation, the MERRA model showed a lower R^2 (0.49) than the TROPOMI (0.57) and OMI (0.54) models, indicating some reduction in spatial generalization associated with the coarser MERRA-2 aerosol fields. Thus, we interpret the similar overall performance as evidence of the importance of regional-scale atmospheric processes, while the satellite-derived predictors provide complementary information for representing spatial variability. We have clarified this interpretation in the revised manuscript.

- Line 692: how are the northern and southern regions of Ghana shown in Figure 3B, or is this a typo?

  Author's response: Figure 3B is now revised to explicitly show the regional grouping: Northern Ghana is represented by solid-filled violins, while Southern Ghana is represented by hatched violins. The season is indicated by color, and a separate "Region" legend has been added for clarity.

  Revision: Figure 3B has been revised accordingly.

- Figure 5: Allowing each subplot to have its own y-axis limit based on its data range or using a consistent one that doesn't extend all the way down to 0 µgm-3 might allow readers to gauge variations a little better, even if it means omitting the WHO guidelines.

  Author's response: We revised Figure 5 to improve the visualization of temporal variations in annual mean $PM_{2.5}$ concentrations for all cities. The revised figure uses a broken y-axis with row-specific upper limits, and the WHO annual guideline has been removed as suggested.

  Revision: Figure 5 and its caption have been updated accordingly.

- Lines 740ff: why are p-values only quoted for some cities? Are p-values even necessary if it was already stated that trends are significant?

  Author's response: We agree that reporting p-values for only some cities was inconsistent and that the individual p-values are not necessary when the sentence starts with mentioning cities with statistically significant trends. We have therefore removed the p-values throughout the sentence and kept only the Sen's slope estimates to describe the magnitude of the trends.

  Revision: The p-values have been removed from the sentence as suggested.

- Paragraph beginning at line 767: When discussing contrasting trends in northern versus southern Ghana, the authors highlight how improved technology e.g., cleaner vehicles, cleaner cooking, etc.) may have reduced PM in southern cities but then point out how northern cities such as Wa and Tamale have positive non-Harmattan trends and suggest that local emissions may be at play in driving the increases. Do the authors have any reason to believe that there is uneven rollout of these technologies in the north versus south or any regional initiatives that would create this dipole by which they would conclude that improved technology only reduces PM in the south?

  Author's response: Our intention in this paragraph was to emphasize that the stronger declines observed in several southern urban centers may reflect the greater degree and spatial extent of urbanization in southern Ghana, where changes in urban-related emissions and household energy use may have a more pronounced influence on ambient $PM_{2.5}$. Additionally, there is evidence of historically more limited access to and adoption of cleaner household fuels such as LPG in northern Ghana, concentrated with rural areas, partly due to differences in fuel availability and distribution infrastructure. We have revised the text to clarify this distinction and to avoid implying that cleaner technologies only reduce $PM_{2.5}$ in southern Ghana.

  Revision: The last paragraph in Section 3.3 is now revised to include the following text.

  "*The stronger declines observed in several southern urban centers may reflect, in part, the greater degree and spatial extent of urbanization in southern Ghana, where changes in urban-related emissions and household energy use may have a more pronounced influence on ambient $PM_{2.5}$. Additionally, although national initiatives have promoted the adoption of cleaner household fuels such as LPG, access to and adoption of LPG have historically been more limited in northern and rural areas, partly due to differences in fuel availability and distribution infrastructure.*[51] *In northern Ghana, the influence of biomass burning and* *other regional sources may partially offset or obscure reductions in urban-related emissions.*"

- Furthermore, in Lines 892-896 the authors stated that declines in southern population centers are consistent with improvements in vehicle fuel standards, emissions controls, and incremental adoption in cleaner household energy. Similar to my question above, is there uneven rollout of these initiatives in the northern cities? Or why wouldn't this argument be used to understand Ghana-wide urban trends?

  Author's response: We agree that improvements in vehicle fuel standards, industrial emissions controls, and adoption of cleaner household energy can affect urban areas across Ghana and are not specific to the southern regions. Our intention was to emphasize the greater degree and spatial extent of urbanization in southern Ghana, where population and urban activities are more concentrated, and therefore reductions in urban-related emissions may have a more pronounced influence on the observed $PM_{2.5}$ trends.

  Revision: We have revised the text in the Conclusion section to clarify this distinction.

  "*The observed declines in several southern urban centers may be consistent with improvements in vehicle fuel standards, industrial emission controls, and incremental adoption of cleaner household energy. These factors may also contribute to urban $PM_{2.5}$ trends in other parts of Ghana, but their influence may be more pronounced in southern Ghana because of the greater degree and spatial extent of urbanization and concentration of urban activities in this region.*"

- Line 833ff: Isn't it the *major* wet season that shows a partial recovery of the north-south gradient? The minor wet season PM plot in Fig. 7E looks somewhat anthropogenic/correlated with population density and biomass burning in Figs. S1, S10.

  Author's response: We understand that the spatial pattern during the wet seasons warrants clarification. The reviewer refers to the minor wet-season panel in Fig. 7, which is panel D in the revised figure. We have revised the text to distinguish the two wet seasons more clearly. Specifically, the major wet season shows a substantial weakening of the north–south gradient, whereas during the minor wet season, $PM_{2.5}$ remains relatively low but exhibits a more heterogeneous spatial pattern, with elevated concentrations extending across parts of central and southern Ghana.

  Revision: We have revised the corresponding text in Section 3.5 accordingly.

  "*During the major wet season (Fig. 7C), $PM_{2.5}$ concentrations drop significantly throughout the country, and the north-south gradient weakens considerably, driven by* *enhanced wet deposition and suppressed dust transport, while southern coastal cities retain discernible $PM_{2.5}$ elevations relative to the rural background, reflecting persistent contributions from traffic, open waste burning, and industrial activity. During the minor wet season (Fig. 7D), $PM_{2.5}$ remains relatively low but exhibits a more heterogeneous spatial pattern, with elevated concentrations extending across parts of central and southern Ghana.*"

Supplement:

- Table S2: Just curious/checking - is Kintampo supposed to be "rural"?

  Author's response: We have clarified the site classifications in Table S2 to better reflect the spatial characteristics of the monitoring network. In particular, the 35 sensors associated with Kintampo were distributed across the broader Kintampo area rather than being clustered within the town, providing coverage across a range of surrounding rural and peri-urban environments. Therefore, we have classified Kintampo as rural/peri-urban and clarified that the site classifications represent the predominant setting of the monitoring area.

  Revision: Table S2 has now been revised to reflect this change.

- Text S1: For Airnote, where are the corrections coefficients obtained from? There doesn't appear to be a BAM in Kumasi and Somanya from which scalings were obtained? I didn't look up the Owusu-Tawiah et al. (2025) reference to see if it's in there, but - even if it is - I think more explanation should be made available here so if anyone wants to reproduce/extend this works, it's clear to understand the assumptions of important equations like these.

  Author's response: We have revised Supplementary Text S1 to provide additional details about the calibration procedure used to derive correction coefficients for the Airnote sensors. Specifically, we clarify that the MLR model was developed using Airnote measurements collected between September and December 2023 while collocated with a Teledyne T640 reference monitor, following the calibration procedure described by Owusu-Tawiah et al. (2025).

  Revision:

  "*Airnote sensors were mainly deployed in Kumasi (urban) and Somanya (peri-urban). Following the calibration procedure described by Owusu-Tawiah et al. (2025)[2], we developed a multiple linear regression (MLR) model using measurements collected between September and December 2023 from Airnotes collocated with a Teledyne T640* *reference monitor. Airnote-measured $PM_{2.5}$, temperature, and relative humidity (RH) were used as predictor variables, while $PM_{2.5}$ measurements from the T640 were used as the response variable. The resulting equation was:*

$$PM_{2.5,\ corrected} = -6.6 + 0.57*PM_{2.5,\ raw} + 0.28*T + 0.03*RH \ \ldots(S2)$$”

- Fig. S3: should the x- and y-axis tick label currently written as “NO” be “NO_2”?

  Author’s response: The axis labels in Figure S3 have been corrected from NO to NO_2.

  Revision: Figure S3 has been revised.