Response to Reviewer #1

Summary

The authors develop a 1 km × 1 km daily surface PM2.5 product for Ghana spanning 2005–2025, using three XGBoost models (OMI-based for 2005–May 2018, TROPOMI-based for May 2018–2025, and a MERRA-2-based gap-filler) trained on a 98-monitor ground network (96 calibrated low-cost sensors + 2 BAMs, Aug 2018–Oct 2024). They evaluate five ML algorithms under random, temporal, and spatial cross-validation; interpret predictors with SHAP; and characterize long-term trends with Seasonal Mann–Kendall and Sen's slope. The headline finding is a persistent, possibly widening north–south exposure disparity. The dataset is publicly released on Zenodo.

This is a valuable contribution. It is the first long-term high-resolution PM2.5 record for Ghana, built on an impressive ground-network assembly and a sensible multi-model design, with commendable data release and appropriate emphasis on spatial CV as the operationally relevant metric. My concerns center not on the modeling framework, which is soundly constructed, but on the extent to which the central *trend* claims are supported given where and when observations actually constrain the model.

Author's response: We appreciate your kind remarks on our work and sharing the comments. We have addressed the concerns through revisions in the main text and the supplementary information.

Recommendation: Minor revision. Here are some of my major concerns:

1. The temporal record is largely extrapolated beyond the training window, and this is underweighted in the framing. Ground data exist only for Aug 2018–Oct 2024, yet PM2.5 is estimated back to 2005. The OMI model is the extreme case: it is trained on 2018–2024 OMI retrievals and applied to 2005–May 2018, so its *entire* application domain is out-of-sample in time. This coincides with the evolution of the OMI row anomaly (which the authors themselves cite via Torres et al., 2018) and with drift in the predictor set's characteristics over two decades. The leave-one-year-out CV only spans 2018–2024 and cannot speak to 2005–2017 performance. Please state explicitly that ~13 of 21 years are unvalidated, and discuss how instrument aging and predictor non-stationarity could bias the retrospective estimates and, critically, the trends.

Author's response: Thank you for your valuable comments. We agree that direct surface validation is limited to August 2018–October 2024 and that the leave-one-year-out temporal cross-validation cannot directly assess model performance before the monitoring period.

OMI retrievals were subjected to product-specific quality screening, including cross-track quality flags and explicit row-anomaly filtering, and temporal leave-one-year-out validation demonstrated the model's ability to generalize across unseen years within the available observation period. We therefore regard 2005–2017 as a temporally extrapolated portion of the dataset rather than directly validated observations and have revised the framing of the long-term estimates and trends accordingly.

Revision:

Section 2.2 (in Materials and methods)

"*$PM_{2.5}$ 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 3.3 (in Results and discussion)

"*Changes in the characteristics of valid satellite retrievals or in their relationship with surface $PM_{2.5}$ over the longer record could introduce systematic differences in the retrospective estimates that cannot be directly evaluated in the absence of pre-2018 surface observations. Such differences could also influence the magnitude of estimated long-term trends and should therefore be considered when interpreting the retrospective record.*"

2. The central finding sits in the region with the least ground constraint. The monitoring network is concentrated in the middle belt (Kintampo, Techiman) and south (Accra, Kumasi); Tamale (~9.4°N) is effectively the only genuinely northern site, with nothing in Upper East, Upper West, North East, or Savannah. The dramatic north–south gradient and the "rising north" signal therefore derive from model extrapolation into unmonitored terrain rather than from measurements. This is sharpest in Fig. 5: of the cities with statistically significant *increasing* trends (Wa +0.11, Nalerigu +0.07, Damongo +0.05, Sunyani +0.04 $\mu g\ m^{-3}\ yr^{-1}$), none has a monitor. The spatial CV cannot even test northern performance because there are no northern sites to withhold. Fig. S13 makes the distance-to-sensor problem clear. I'd ask the authors to (a) explicitly delimit where the product is observationally constrained vs. extrapolated, and (b) soften the abstract and conclusions accordingly; the divergence claim should be presented as suggestive given the monitoring geometry.

Author's response: We agree that the monitoring network is spatially heterogeneous and that uncertainty may be higher in areas that are farther from monitoring locations. We initially highlighted this in the manuscript through Figure S15. The Materials and Methods section now clarifies that monitoring in northern Ghana is limited to Tamale, where three QuantAQ PM monitors collected $PM_{2.5}$ measurements over approximately one year and were included in the spatial cross-validation framework, providing some direct observational constraint within the northern belt.

We have also made the following changes in the main text to account for this limitation while interpreting the trend results.

Revision: The sentence in the Abstract now uses "suggests" instead of "reveals" to present the north-south divergent claim as a possibility instead of a definitive claim.

"*The long-term gridded $PM_{2.5}$ dataset suggests a unique north-south exposure disparity in Ghana, with northern regions of the country experiencing substantially higher $PM_{2.5}$ concentrations compared to the South, that may be widening over the 21-year period by over 0.2 µg m^{-3} yr^{-1}.*"

In the Materials and Methods section, we have highlighted that monitor density in the north is lower compared to central and southern Ghana.

"*The monitoring network had broader coverage across central and southern Ghana, with 35 sensors in the Kintampo area, 16 at Techiman, 20 at Accra, and 13 at Kumasi, while northern Ghana was represented by three sensors in Tamale (Table S2; Fig. S1). The sensors associated with Kintampo were distributed across the broader area surrounding Kintampo rather than being concentrated solely within the Kintampo town.*"

In the Results and Discussions section, we revised the trends paragraph as:

"*Overall, long-term changes in $PM_{2.5}$ across Ghanaian suburban and urban locations are modest, typically ranging between -0.2 µg m^{-3} yr^{-1} and 0.2 µg m^{-3} yr^{-1}. These trends suggest a persistent north–south gradient in $PM_{2.5}$ exposure over the 21-year period, with geographically divergent long-term trends superimposed on substantial interannual and multi-year variability. Positive trends in northern Ghana contrast with stable or declining trends in the south, reflecting differences in dominant drivers of $PM_{2.5}$ variability across regions. This north–south pattern should be interpreted in the context of the spatial distribution of monitoring sites, particularly the lower monitoring density in northern Ghana.*"

In the Conclusion section, we have highlighted this limitation.

"*Spatial cross-validation performance ($r \approx 0.73$–0.76) indicates that the XGBoost models capture dominant spatial patterns across Ghana, though prediction uncertainty may be higher in areas far from monitoring sites or emission environments not well represented by the monitoring network, particularly in the sparsely monitored northern regions of Ghana (Fig. S15).*"

3. Continuity across the OMI→TROPOMI→MERRA seams is not demonstrated, and the blending scheme is not described. Three models with different predictor spaces and different mean biases (Table 1) are merged into one "spatiotemporally complete" product, with a hard temporal seam at May 2018 and spatially/temporally variable seams wherever gap-fill activates. No method is given for how the models are combined, and no consistency check is shown.

Author's response: We have now clarified that the final product is generated using a hierarchical prediction and gap-filling procedure, where the OMI model is used from January 1, 2005 to May 6, 2018, the TROPOMI model from May 7, 2018 onward, and the MERRA-2 model is used only for grid cells and days where the corresponding satellite-based prediction is unavailable. We have explained this model-selection procedure in the revised Section 2.5 as:

"*For each day and 1-km grid cell, the corresponding satellite-based model (TROPOMI or OMI) was used when the required satellite predictors were available. Grid cells where a satellite-based prediction could not be generated were then filled using the MERRA-2-based model. Thus, the final $PM_{2.5}$ product was constructed through a hierarchical prediction and gap-filling procedure. The OMI model was used from January 1, 2005 to May 6, 2018, and the TROPOMI model from May 7, 2018 onward, with the MERRA-2 model providing predictions only for remaining spatial or temporal gaps.*"

As a consistency check, we examined the annual and seasonal national mean $PM_{2.5}$ time series across the OMI-to-TROPOMI transition shown in Fig. 4A. The transition does not show any discontinuity or sudden change in either the annual or seasonal mean $PM_{2.5}$ records. The 2018 transition year, during which the OMI-based model with MERRA-2 gap filling was used through May 6 and the TROPOMI-based model with MERRA-2 gap filling thereafter, follows the temporal progression between the 2017 OMI + MERRA-2 and 2019 TROPOMI + MERRA-2 estimates. This consistency across the annual and seasonal averages provides additional confidence that there is no pronounced discontinuity associated with the transition between the satellite-based model configurations.

4. The authors' own supplementary figures undercut the monotonic "widening" narrative. The successive decadal Harmattan-season deltas are $\Delta = -2.9$ (S16), +4.0 (S17), −0.9 $\mu g\ m^{-3}$ (S18),i.e., the net +2.1 in Fig. 8 is dominated by a single elevated 2015–2019 window, not a secular increase, and the sign is not consistent across sub-periods. This is exactly the signature of multi-year dust variability rather than a trend, and it means the two-half comparison in Fig. 8 is highly sensitive to window choice. Please reconcile the main-text narrative with Figs. S16–S18 explicitly, and consider a robustness check (e.g., trend sensitivity to excluding extreme dust years, or to alternative window definitions).

Author's response: We acknowledge the substantial variability in the north–south $PM_{2.5}$ contrast across different time periods. We agree that the magnitude of the north–south difference is not monotonic across all sub-periods and that the two-period comparison in Figure 8 is therefore sensitive to the selected time windows. We have revised the text to distinguish the persistent north–south spatial gradient and the geographically divergent long-term trends identified from the spatially resolved trend analysis (Fig. 4B) from a uniformly increasing magnitude of the north–south concentration difference over time. The period-by-period comparisons in Figs. S18–S20 (previously S16–S18) further demonstrate the substantial interannual and multi-year variability underlying these spatial patterns.

Revision: We have revised the texts as described below.

Section 3.3: trend interpretation

"*Overall, long-term changes in $PM_{2.5}$ across Ghanaian suburban and urban locations are modest, typically ranging between -0.2 µg m^{-3} yr^{-1} and 0.2 µg m^{-3} yr^{-1}. These trends suggest a persistent north–south gradient in $PM_{2.5}$ exposure over the 21-year period, with geographically divergent long-term trends superimposed on substantial interannual and multi-year variability.*"

Section 3.3: Fig. 8 discussion

"*In this inter-decadal comparison (Fig. 8), the most notable increases are concentrated along the northern belt, while decreases are evident in the southeastern coastal zone. Period-by-period comparisons at finer temporal resolution, comparing 2005–2009 with 2010–2014, 2010–2014 with 2015–2019, and 2015–2019 with 2020–2025, are provided in Figs. S18–S20. Together, these analyses document persistent north–south differences in $PM_{2.5}$ exposure over the 21-year record, with contrasting long-term trends between a dust-burdened north and a south showing modest but measurable improvements, superimposed on substantial interannual and multi-year variability.*"

5. Trend significance is asserted more strongly than the aggregate statistics support. While the 2025 Harmattan PM2.5 increase and other sub-wet time remain consistent, the annual average decrease? Are you using the population-weighted mean PM2.5 or pixel average PM2.5? The significance of the reported trends is asserted more strongly than the aggregate statistics support. The Ghana-wide annual and seasonal Sen's slopes are non- significant (Fig. 4A; annual $p = 0.408$, lines 717–721), yet the pixel-wise analysis is described as significant across "the majority of Ghana" (Fig. 4B, lines 721–723). Pixel-wise Mann–Kendall (method described in §2.7.1, lines 456–470) applied to ~10^5 spatially autocorrelated grid cells, without a field-significance or multiple-comparisons correction, will flag a large number of false positives, and spatial autocorrelation causes these to cluster into apparently coherent regions. Please apply an appropriate correction (e.g., Benjamini–Hochberg FDR, or adjustment for the effective number of independent cells) and re-interpret Fig. 4B (lines 721–723) accordingly. Relatedly, the argument that annual averaging attenuates the 7–21 µg m^{-3} daily RMSE (lines 747–755) holds only for unsystematic error; time-varying *systematic* error (instrument drift, seam offsets, changing gap-fill fraction) survives temporal averaging and could alias into the trend estimates. Please confirm whether such systematic effects can be excluded.

Author's response: The positive Harmattan-season trend can coexist with a modest decline in the annual mean because the annual mean integrates $PM_{2.5}$ across all seasons, and changes during the Harmattan season can be offset by $PM_{2.5}$ behavior during the remainder of the year. The Ghana-wide and regional time series were calculated as spatial averages of the 1-km grid-cell $PM_{2.5}$ concentrations and were not population-weighted. This has been clarified in Section 2.7.

"*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.*"

We evaluated the spatial autocorrelation of our $PM_{2.5}$ fields at 0.01° × 0.01° grids using Global Moran's I and compared the results with the established gridded $PM_{2.5}$ product of Shen et al. (2024) at both 0.01° and 0.1° spatial resolutions. All products exhibited strong positive spatial autocorrelation, supporting the interpretation that spatial dependence is an inherent characteristic of gridded $PM_{2.5}$ fields. The results are presented in Text S3 and Table S12.

We also applied the Benjamini–Hochberg false discovery rate (FDR) correction to the pixel-wise Mann–Kendall p-values to account for multiple comparisons across the spatial domain. Of the 195,811 valid 1-km grid cells, 137,333 (70.14%) were significant using the uncorrected $p < 0.05$ criterion, while 131,061 (66.93%) remained significant after FDR correction at $q < 0.05$. Thus, 95.43% of the originally significant pixels remained significant after correction. The Sen's slope estimates were unchanged and only the significance classification was adjusted. The revised Fig. 4B now shows the FDR-corrected significant trends, and the FDR analysis is described in Text S4 and outputs are summarized in Table S13.

We agree that temporal averaging does not eliminate time-varying systematic errors. We therefore revised the text to clarify that annual averaging is expected primarily to attenuate unsystematic daily errors. As an additional assessment of temporal robustness, we performed leave-one-year-out cross-validation for 2018–2024, evaluating the model's ability to generalize across interannual variability in meteorology, emissions, and episodic pollution events. The revised text says:

"*However, the trend detection analysis was performed on annual means, for which unsystematic daily errors are expected to be attenuated through temporal averaging, reducing their contribution to the uncertainty of the annual estimates. To further assess temporal robustness, we evaluated model performance using leave-one-year-out cross-validation for 2018–2024, testing model generalization across interannual variability in meteorology, emissions, and episodic pollution events.*"

6. Sensor-type calibration is confounded with location, and calibration uncertainty is not propagated. Sensor types are spatially clustered (PurpleAir entirely in Techiman, QuantAQ dominated by Kintampo, Clarity dominated by Accra), and each type carries a distinct calibration; the Clarity model in particular (intercept 54.6, raw slope 0.4) is an aggressive correction. Any spatial pattern in the "ground truth" therefore partly reflects between-type calibration differences rather than true PM2.5 gradients, and none of the calibration uncertainty enters the reported model or trend uncertainty. At minimum, please (a) discuss this confound and its potential contribution to the north–south contrast, and (b) comment on how calibration error would propagate into the trend estimates.

Author's response: We agree that the sensor types are not uniformly distributed across Ghana, but this spatial structure partly reflects an intentional design of the monitoring network. Sensors were intentionally deployed at higher densities in more populated urban areas to capture intra-city spatial variability associated with diverse emission sources and high population density, while additional monitors were deployed in suburban and rural areas to capture broader geographic and environmental variability. We have clarified this network design in Section 2.2 (revised text below). In addition, the 35 sensors in the Kintampo area are distributed across the broader area surrounding Kintampo rather than being concentrated solely within Kintampo town (Fig. S1).

To further assess whether this spatial clustering could contribute to differences among sensor types, an independent intercomparison of QuantAQ Modulair-PM, PurpleAir PA-II, and Clarity Node-S monitors conducted in Accra, Ghana, found comparable post-correction performance across the three sensor types (Raheja et al., 2023). In their best-performing XGBoost correction models, the CvMAE was 0.03, 0.05, and 0.03 for PurpleAir, Clarity, and Modulair-PM, respectively. This provides evidence that, following calibration, differences among these sensor types do not necessarily translate into substantial differences in measurement error. The relatively large intercept in the Clarity calibration equation therefore does not, by itself, indicate greater calibration uncertainty. The raw sensor measurements were not used directly as ground observations, and all low-cost sensor measurements were first subjected to sensor-specific calibration and QA/QC before model development.

Regarding propagation to the long-term trends, the calibration equations are linear functions of the raw $PM_{2.5}$ measurement and meteorological variables (ambient temperature and relative humidity). An additive calibration intercept does not affect the Sen's slope because adding a constant does not change the pairwise slopes used to estimate the trend. In contrast, uncertainty in a multiplicative calibration coefficient can scale the magnitude of the estimated trend proportionally. Because the calibration equations also include temperature and relative humidity terms, time-varying uncertainty associated with these predictors can affect the calibrated concentrations differently from a simple additive offset.

Revision: Section 2.2 is revised as:

These sensors were strategically deployed across Ghana, with a significantly higher density in urban areas where population is largely concentrated, to capture high spatial variability in $PM_{2.5}$ concentrations arising from diverse sources, including vehicular emissions, open burning, cooking activities and industrial emissions. Air quality monitors were also deployed in suburban and rural regions of Ghana, where local air quality is strongly influenced by emissions from solid biomass fuel use for cooking, agricultural residue burning and dust from unpaved roads. This sensor network architecture helps inform the model about the spatial distribution of $PM_{2.5}$ in Ghana during the Harmattan period by using measurements from northern, middle and southern belts of

Ghana. The monitoring network had broader coverage across central and southern Ghana, with 35 sensors in the Kintampo area, 16 at Techiman, 20 at Accra, and 13 at Kumasi, while northern Ghana was represented by three sensors in Tamale (Table S2; Fig. S1). The sensors associated with Kintampo were distributed across the broader area surrounding Kintampo rather than being concentrated solely within the Kintampo town. Fig. S1 depicts the locations of every monitor against a population count map retrieved from the LandScan project at Oak Ridge National Laboratory.[28] Additional deployment details are available in Tables S2–S3.

Raw $PM_{2.5}$ measurements from each low-cost sensor, available at sub-hourly time resolution, were aggregated to hourly concentrations and subjected to custom-designed calibration models unique to each sensor type to derive QA/QC-corrected exposure levels. The calibration procedures are further explained in Supplementary Text S1. These sensor-specific corrections were applied prior to model development to improve the consistency of $PM_{2.5}$ measurements across the different sensor types.

Minor / specific comments

1. "GRASP" appears only in the Fig. S2 caption and is never defined. Define it or remove.

Author's response: The word "GRASP" has been removed from the Fig. S2 caption.

2. Author-name inconsistencies: "Clement Mensah Ackaah" (main) vs. "Clement Ackaah (SI). Reconcile.

Author's response: The name of the author is now corrected to "Clement Mensah Ackaah" in the SI as well.

References

Raheja, A., et al. (2023). *Low-Cost Sensor Performance Intercomparison, Correction Factor Development, and 2+ Years of Ambient $PM_{2.5}$ Monitoring in Accra, Ghana*. Environmental Science & Technology, 57, 10708–10720. DOI: 10.1021/acs.est.2c09264.