Summary

The authors develop a 1 km × 1 km daily surface $PM_{2.5}$ 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 $PM_{2.5}$ 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.

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 $PM_{2.5}$ 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.
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.
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.

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 µg 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).
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 $\sim 10^5$ 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.
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 $PM_{2.5}$ 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.

Minor / specific comments

- "GRASP" appears only in the Fig. S2 caption and is never defined. Define it or remove.
- Author-name inconsistencies： "Clement Mensah Ackaah" (main) vs. "Clement Ackaah" (SI). Reconcile.