Title: Enhancing Low-Cost $PM_{2.5}$ Sensor Reliability Through Multi-Model Calibration Against a Beta Attenuation Monitor

I appreciate the opportunity to review this manuscript. The topic is important and relevant for the community: improving low-cost $PM_{2.5}$ sensing performance is valuable for broader air-quality monitoring applications. However, after assessment, I recommend rejection in its current form because the manuscript does not yet establish a clear novel contribution and requires substantial reworking throughout.

The manuscript reports calibration of a low-cost $PM_{2.5}$ sensor against BAM data using multiple statistical/ML models. While performance improvements are reported, the current study has foundational issues in novelty framing, contextual reporting, analytical rigor, and presentation. In my view, the required changes are not incremental; they involve redesign of the framing and substantial redevelopment of the analysis.

Major Concerns

Section1 Introduction

The manuscript frames a key gap around humidity/temperature effects on low-cost optical PM sensor response. This relationship is already well documented in existing literature, and the authors summarise some literature in Table 1. The current manuscript does not clearly demonstrate what is new beyond known behaviour and commonly used calibration workflows. A publishable revision would need a substantially stronger, explicit novelty claim (e.g., transferability, understudied geographical region, or clearly differentiated methodological advance) and direct evidence supporting that claim.

Response to Reviewer 1 – Section 1 (Introduction)

We sincerely thank the reviewer for the thoughtful and detailed assessment. We acknowledge the concern regarding novelty framing and agree that the original manuscript did not sufficiently articulate its distinct contribution relative to existing calibration studies.

While humidity influence on optical PM sensors is well documented, our contribution extends beyond demonstrating environmental bias. The revised manuscript clarifies that the novelty of this work lies in:

1. **A structured multi-model comparative calibration framework** applied under tropical high-humidity conditions (up to 95% RH), where sensor over-response behavior is extreme and less studied compared to temperate regions.
2. **A hybrid adaptive-blend calibration model**, integrating physics-based linear correction with nonlinear ensemble learning, designed to balance interpretability and predictive robustness.
3. **Independent validation using subsequent co-location datasets**, ensuring that calibration performance is not limited to in-sample fitting.
4. **Comprehensive agreement analysis**, including Bland–Altman limits of agreement, proportional bias assessment, and environmental-stratified evaluation.

We have substantially revised the Introduction to:

- Explicitly define the research gap,
- Differentiate this work from prior humidity-aware calibration studies,
- Clarify the methodological advancement,
- Emphasize regional and climatic relevance.

Additionally, the Methods and Results sections have been restructured to improve analytical rigor, including clearer training–validation separation and enhanced justification of model selection.

We appreciate the reviewer's guidance, which has significantly strengthened the manuscript.

Section 2 Methods

The Methods section contains excessive low-level electronics/build detail in the main text, which disrupts the scientific narrative and reads more like thesis documentation. The manuscript should be restructured so core measurement/calibration methods are the focus, with implementation-level hardware details moved to Supplementary Material or a repository.

Key experimental information is missing from the methods.

Sensors used

In the text (line 118) three sensors are mentioned (GP2Y, ZH03 and SDS011) to be used in the study and all are included in Table 2. However, only data from ZH03 is included in the analysis and reported on.

Outlier removal

The authors report removing outliers using the Z-score method and in text (line 260) say they removed isolated concentration spikes which were not erroneous data. High concentration events are arguably more important in air pollution contexts due to the relationship between high concentrations and negative health impacts. No information was provided on the amount of data which were removed, no sensitivity analysis was completed to assess impact on the model performance. It would also be important to know if the removed events were in the training or test data.

Site information/Experimental Approach

There is no information on the site of the study, where did the co-location take place? What site type classification would be appropriate? (e.g. roadside, kerbside, background etc). There are no start/end dates for the co-location period reported in the methods. Lacking seasonal/geographical context.

Lack of train/test splitting information and hyperparameter searching approach.

There is insufficient clarity in the modelling descriptions as no dependent variables explicitly mentioned. There is a reference to a physics-based linear correction for the blended model (line 317), but this is not described before and appears to be the previous linear scaling.

Response to Reviewer 1 – Section 2 (Methods)

We thank the reviewer for the detailed and constructive comments. We agree that the Methods section required restructuring and clarification. The manuscript has been revised accordingly as follows:

1. **Hardware Detail Reduction**

   Low-level electronics and implementation details (power supply, wiring architecture, communication modules) have been moved to the Supplementary Material. The main text now focuses exclusively on measurement configuration, co-location design, data preprocessing, calibration framework, and validation strategy.

2. **Sensor Clarification**

   Although three low-cost PM sensors (GP2Y, ZH03, SDS011) were introduced in Table 2 for comparative specification purposes, only the ZH03A sensor was used in the calibration analysis. The text has been revised to clarify this explicitly and avoid ambiguity.

3. **Outlier Removal Transparency**

   The manuscript now reports:

   - The percentage of data removed during outlier filtering.
   - A sensitivity analysis comparing model performance with and without outlier removal.
   - Clarification that removal was applied only to the training dataset and not to the independent test dataset.

4. **Site and Experimental Context**

   A new subsection has been added describing:

   - Study location and geographic coordinates.
   - Site classification (e.g., urban background).
   - Start and end dates of co-location.
   - Climatic context.

5. **Model Training and Validation Details**

   The revised manuscript explicitly describes:

   - Chronological train–test splitting strategy.
   - Hyperparameter tuning procedure.
   - Dependent and independent variables used in each model.
   - Clarification of the physics-based linear correction used in the adaptive-blend model.

We appreciate the reviewer's guidance, which has significantly improved methodological transparency and reproducibility.

Section 3 Results & Discussion

The analysis completed lacks focus on the reported novelty and could be improved for clarity of interpretation. There appears to be no discussion to wider literature and limited discussion on the reported results. It's mostly a results section.

Section 3.1

There is extensive reporting of raw values and summaries of the raw values relative to temperature and humidity. The manuscript also highlights some example periods in Figure 7. This takes up significant space within the section that would be better focused on assessing the calibration models and deeper analysis on the effects of the environmental variables using modelling approaches.

The assessment of the raw data presented in Table 3 is difficult to interpret. A different approach by reporting the difference between the two would help. Inclusion of the number of values with the stratification would be informative also.

From eye there may be some kind of temporal misalignment as the sensor appears to respond an hour before the BAM in figure 7.

Section 3.3

This section looks at correlations to environmental variables (temperature and humidity) to the raw readings of the sensor and the BAM. I am unsure of the relevance of this approach. A more meaningful assessment would be to look at the bias on the y and the environmental variable on the x.

Section 3.4

I would like to see some time series of the calibrated data during these example periods or the data coloured by humidity/temperature, for example. Overall, I would like to see a more rigorous assessment of the calibrated models on temperature and humidity effects, which is the novelty claim of the manuscript. Such as during the highlighted extreme periods in Figure 7, if the train/test split allows. There is the reported overall MAE in Figure 12. From eye there is no meaningful difference between the Random Forest, Gradient Boosted, SVR and Adaptive-blend models. A difference of 2 $\mu g\ m^{-3}$ MAE isof note from the linear model compared to the more computationally intensive ML approach. Claims of substantial improvement are not demonstrated in this manuscript. In addition, the differences in the reported metrics between the ML models are very small. There is no discussion on these points in the manuscript.

Response to Reviewer 1 – Section 3 Results & Discussion

We thank the reviewer for the detailed evaluation of the Results section. We acknowledge that the original manuscript emphasized descriptive reporting of raw measurements and did not sufficiently focus on calibration performance and environmental compensation, which represent the core contribution of the study.

The Results section has been substantially revised to:

1. Reduce descriptive reporting of raw data and move supplementary summaries to the Supplementary Material.
2. Reframe Section 3.1 to focus on bias characterization rather than raw value summaries.
3. Replace simple correlation analysis with bias-versus-environment analysis.
4. Include additional time-series visualizations of calibrated outputs during extreme humidity episodes.
5. Add stratified performance evaluation across humidity regimes.
6. Provide clearer comparison among calibration models and discuss the practical significance of performance differences.
7. Expand discussion with reference to relevant calibration literature.

These revisions improve clarity, strengthen the novelty positioning, and align the Results section with the stated objectives of environmental compensation and hybrid calibration.

Section 4 Conclusion and Future works

I am not convinced that the conclusions drawn are compatible with the reported analysis/experimental approach.

Line 588 says "*a comprehensive calibration framework was developed*'. No definition of framework was outlined in this manuscript and from my understanding of a framework, this was not demonstrated. Here, a calibration report has been presented.

The post calibration time series presented in Figure 13 is for a two-day period. It is not long enough to describe the data as showing measurements across varied pollution episodes and meteorological conditions. This was not shown in the data.

Response to Reviewer 1 – Section 4 Conclusion and Future Work

We thank the reviewer for the careful evaluation of the conclusions and for highlighting areas where the manuscript overstated its claims. We agree that the terminology and interpretation required clarification. The manuscript has been revised accordingly.

First, the wording describing the proposed method as a "comprehensive calibration framework" has been revised. In the original submission, the term "framework" was intended to describe the structured combination of linear correction, nonlinear machine-learning calibration, and adaptive blending. However, we acknowledge that the study does not implement a full modular calibration architecture in the sense commonly associated with calibration frameworks. The text has therefore been revised to describe the contribution more accurately as a **hybrid calibration approach** or **multi-model calibration strategy** rather than a framework.

Second, we acknowledge the reviewer's comment regarding the limited duration of the post-calibration time series presented in Figure 13. The two-day example was originally included to visually illustrate the behavior of calibrated sensor outputs relative to the BAM reference during representative periods. However, this short window does not constitute sufficient evidence for evaluating performance across diverse pollution episodes or meteorological regimes. The manuscript has therefore been revised to clarify that the figure serves only as an illustrative example, while the main evaluation of calibration performance is based on the full dataset and statistical metrics reported in the Results section.

To improve clarity, the revised conclusion now focuses on the demonstrated findings of the study, namely:

(i) the quantification of humidity-related bias in the raw optical sensor measurements,
(ii) the comparative evaluation of linear and nonlinear calibration approaches, and
(iii) the improved agreement achieved through hybrid calibration under high-humidity conditions.

These revisions ensure that the conclusions are consistent with the scope of the analysis presented in the manuscript.

General Comments

There are formatting inconsistencies throughout the manuscript regarding symbols/subscripts.

Lack of information in table and figure captions to be stand alone.

English and grammar assessment is needed throughout.

There appears to be generative AI used in the text production and for generating plots. This needs to be acknowledged as per journal guidelines.

There is a lack of discussion, therefore relevant literature is not included.

Response to Reviewer 1 (General Comments)

We thank the reviewer for the careful reading of the manuscript and for the constructive comments regarding formatting, clarity, and presentation. The manuscript has been thoroughly revised to address these concerns. The specific changes are described below.

Formatting inconsistencies regarding symbols and subscripts

We acknowledge that the original manuscript contained several formatting inconsistencies in the representation of symbols, subscripts, and mathematical notation. These issues have been corrected throughout the manuscript to ensure consistent formatting of variables, units, and subscripts (e.g., $PM_{2.5}$, PM_{raw}, PM_{ref}). Mathematical expressions have also been standardized using consistent notation across the Methods and Results sections.

Table and figure captions

We agree that several table and figure captions lacked sufficient detail to be fully interpretable as standalone elements. All captions have been revised to include clearer descriptions of the data shown, the variables represented, and the context required for interpretation without referring back to the main text. Where applicable, units, sample sizes, and definitions of abbreviations have also been added.

English and grammar

The manuscript has undergone a comprehensive language revision to improve clarity, grammar, and overall readability. Sentence structures have been simplified where necessary, and redundant wording has been removed. The revised version has been carefully edited to ensure consistent academic writing style throughout the manuscript.

Use of generative AI

We appreciate the reviewer highlighting this point. Generative AI tools were used only as writing assistance to improve language clarity and structure. The scientific content, analysis, and interpretation were developed entirely by the authors. A statement acknowledging the use of AI-assisted writing tools has been added to the manuscript in accordance with the journal's guidelines.

Limited discussion and missing literature

We acknowledge that the original manuscript contained limited discussion of the results in relation to the existing literature. The Results and Discussion section has been substantially expanded to include comparison with previous studies on low-cost PM sensor calibration and environmental compensation approaches. Additional references have been incorporated to better contextualize the findings and highlight similarities and differences with previously reported calibration performance.

These revisions have improved the clarity, consistency, and contextualization of the manuscript.

Closing remarks

Although the topic is relevant, the current manuscript does not provide a sufficiently clear or well-supported novel contribution, and the analysis requires substantial redevelopment rather than routine revision. I therefore recommend rejection at this stage. I encourage the authors to consider preparing a substantially redesigned manuscript with a clearly defined novelty, fuller contextual reporting, and deeper scientific discussion.

Response to Reviewer 1 (Closing Remarks)

We thank the reviewer for the careful evaluation of the manuscript and for the constructive feedback provided throughout the review. We appreciate the reviewer's concerns regarding the clarity of the novelty, the depth of analysis, and the contextual discussion presented in the original submission.

In response, the manuscript has been **substantially revised and reorganized** to address these concerns. In particular, we have implemented the following major improvements:

1. Clarification of the study contribution
   The Introduction and Methods sections have been revised to clearly articulate the contribution of this work as a **hybrid calibration approach combining linear**

correction and nonlinear machine-learning models, evaluated under high-humidity tropical conditions. The novelty statement has been refined to avoid overstating the contribution.

2. Restructuring of the Results and Discussion section
   The Results section has been reorganized to focus more directly on calibration model performance and environmental bias characterization. Descriptive reporting of raw measurements has been reduced, while additional analysis of **sensor bias relative to temperature and relative humidity** has been included.

3. Expanded discussion and literature comparison
   The manuscript now includes a more comprehensive discussion of the results in relation to existing studies on low-cost $PM_{2.5}$ sensor calibration and environmental compensation. Relevant literature has been added to better contextualize the findings.

4. Improved methodological transparency
   Additional details regarding the experimental setup, train–test splitting, outlier handling, and model validation procedures have been incorporated to strengthen the reproducibility and clarity of the analysis.

5. Revisions to conclusions and claims
   The Conclusion section has been rewritten to ensure that the interpretations and claims are fully supported by the presented analysis.

We believe that these revisions have significantly strengthened the manuscript by improving clarity, analytical rigor, and contextual discussion. We are grateful for the reviewer's detailed feedback, which has been instrumental in guiding these improvements.