We sincerely thank Reviewer #2 for their thorough evaluation of our manuscript and for their constructive, insightful feedback. The initial submission had several shortcomings and oversights, and we are truly grateful for the time and care the reviewer invested in identifying these issues. Their detailed guidance has helped us significantly improve the clarity and quality of the paper. We have addressed every comment point-by-point below, with Reviewer #2's comments in **bold** and our responses in blue.

The authors seek to improve the low-concentration performance of a commercial beta attenuation monitor (Thermo 5014i) through two approaches: (1) reducing the particle deposition area (2.0 to 0.4 cm²) to increase mass loading density, and (2) simulating a multi-unit sensor array by partitioning a single instrument's continuous data record into sequential segments and ensemble-averaging those segments. The study also evaluates moving-average smoothing windows to further reduce measurement variability. The authors report reducing the response time from 60 to 22 min while achieving a reported detection limit of 7.75 µg m⁻³.

The central premise of the manuscript is that averaging sequential measurements from a single beta source and detector is representative of averaging measurements from multiple independent source-detector units. This assumption is fundamental to the study's conclusions and requires quantitative justification. Specifically, the authors should explain why sequential segments acquired from the same instrument constitute a valid surrogate for independent physical units and provide evidence that non-Poisson sources of error, such as instrumental drift, detector stability, electronic noise, source variability, tape positioning, or environmental effects, are negligible over the effective integration times considered. Without such evidence, it is unclear whether the observed improvement reflects only the reduction of random counting error or can be generalized to an ensemble of independent BAM instruments.

The manuscript has fundamental conceptual and methodological problems that undermine its central claims, along with substantial internal inconsistencies in the reported data. I detail these below.

The most significant issue:

The central motivation of this manuscript is explicitly framed around ultra-low-concentration $PM_{2.5}$ monitoring. The authors cite the 2021 WHO Air Quality Guideline revision, which lowered the annual $PM_{2.5}$ level to 5 µg m⁻³, and argue that ambient concentrations are increasingly approaching the operational detection limits of commercial BAMs. The stated objective of the study is therefore to enhance BAM performance under these low-concentration conditions. However, the results presented do not appear to support this central motivation.

The manuscript identifies the integrated configuration of Ts + A + N (30-min smoothing, 0.4 cm² filter area, and six-unit ensemble) as the "optimal" configuration. However, this configuration achieves a detection limit (LOD) of 7.75 μg m⁻³ (Table 3, abstract, and conclusions), which remains above the 5 μg m⁻³ WHO guideline that motivates the study. Moreover, this LOD is essentially comparable to, and slightly higher than, the baseline BAM LOD of 8.10 μg m⁻³. Thus, based on the authors' own results, the recommended integrated configuration appears to improve response time but does not provide a meaningful improvement in detection capability relative to the unmodified instrument. It therefore does not resolve the low-concentration measurement challenge that forms the basis of the manuscript.

This discrepancy directly affects the manuscript's primary claims. The abstract and conclusions state that the integrated approach satisfies the requirements for ultra-low-concentration monitoring and enables measurements at levels associated with increasingly stringent global air quality standards. However, the reported LOD of 7.75 μg m⁻³ does not demonstrate this capability. Instead, the results suggest that the proposed configuration remains unable to reliably quantify $PM_{2.5}$ concentrations at or below the 5 μg m⁻³ level used to justify the work.

We acknowledge that the term 'optimal' as applied to the "Ts+A+N" configuration was not sufficiently contextualized in the original draft, leading to a perceived discrepancy regarding our ability to meet the 5 μg m⁻³ WHO Air Quality Guideline.

We have thoroughly revised our discussion to clarify that the 'optimal' configuration represents a balanced engineering compromise designed specifically for capturing high-frequency, rapid spatiotemporal transients rather than an absolute limit of the system's performance. The capability of our proposed framework to resolve the low-concentration measurement challenge is demonstrated through two distinct operational modes:

(1) Equal Temporal Resolution Comparison (Speed-Optimized Mode): If a standard commercial BAM is operated with a 30-minute smoothing window to capture rapid concentration changes, its LOD skyrockets to an unusable 32.2 μg m⁻³ due to elevated counting noise. Under this exact same temporal resolution (22-minute response time), our integrated T_s+A+N configuration suppresses the LOD to 8.28 μg m⁻³. This demonstrates that for rapid-response applications, the integrated strategy drastically enhances detection capabilities over the unmodified instrument.

(2) Ultra-Low-Concentration Mode (Sensitivity-Optimized Mode): Our unified framework is highly adaptable; the temporal smoothing window Ts is a user-defined operational parameter that can be adjusted depending on ambient

concentration levels. If the primary objective is compliance monitoring in clean environments where concentrations are below 5 μg m^{-3}, the smoothing window can be extended. As documented in Table 3 (Column A+N), operating the 0.4 cm^2, 6-unit ensemble with a 75-minute smoothing window suppresses the instrument noise floor to a remarkably low LOD of 1.64 μg m^{-3}. This represents a 83.6% reduction in LOD compared to the reference BAM baseline (10.02 μg m^{-3}) and is well within the required sensitivity to reliably quantify ambient concentrations at or below the 5 μg m^{-3} WHO guideline.

We have updated the Abstract (line 20 – 24), Discussion (line 333-340) , and Conclusions (385-395) to clearly distinguish between these operational profiles, ensuring our claims regarding ultra-low-concentration monitoring are fully supported by our empirical data.

Other major issues:

1. The core novelty of the paper rests on simulating six independent BAM units by splitting one instrument's time series into segments and averaging them. This does not validly emulate a physical array, and the paper never justifies why it should. Averaging N sequential segments from a single source reduces variance by √N only if the segments are statistically independent and identically distributed. But sequential segments from the same physical detector share the same systematic components: the same beta source, the same detector drift, the same filter-tape geometry, the same electronic baseline. A real array of six physical detectors would have six different sources with independent systematic offsets, the very thing this simulation cannot reproduce. What the authors have actually demonstrated is little more than the textbook √N reduction of Poisson counting noise by increasing total integration time. Averaging six one-hour segments is, to first order, statistically equivalent to one six-hour measurement. The paper needs to explicitly address why this is not simply a re-labelling of longer averaging. This distinction is decisive for the whole paper. If the "array" is just longer counting, then the headline conclusions about "array-type beta gauges" and "4–6 units" are not actually supported by the data presented.

We agree that, mathematically, averaging six one-hour segments is statistically equivalent to one six-hour measurement in terms of total Poisson count accumulation. However, we argue that the operational objective of these two approaches is fundamentally different, which is the core premise of our study.

However, we believe the multi-sensor simulation in this study is valid for the following three reasons:

(1) Integration Time vs. Temporal Resolution: The "textbook $\sqrt{N}$ reduction" achieved by a single detector through a six-hour measurement comes at the cost of temporal resolution. A six-hour measurement provides only one data point every six hours, which is insufficient for capturing the rapid concentration transients or traffic-related emission peaks. In contrast, an array of six physical units (simulated by our segments) provides a measurement with that same "six-hour precision" every hour (or less).

(2) Increasing "N" Within a Fixed Time Window: The core of this study is not simply to "count longer," but to increase the number of samples (N) within a fixed, short time window. By combining deposition area reduction with a multi-sensor ensemble, we demonstrate that the system can achieve a 22-minute response time while maintaining the LOD that standard instruments typically require 60 minutes or longer to reach.

(3) Theoretical Architecture vs. Hardware Simulation: We utilized the partitioned dataset from a single BAM to establish the theoretical performance ceiling for an array-type hardware architecture. This simulation allowed us to prove that increasing the spatial sample size is a viable engineering path to breaking the "hourly integration barrier" inherent in single-detector designs. As the reviewer notes, a physical array will face systematic offsets; however, we believe these are manageable engineering challenges (e.g., via pre-operational cross-calibration) rather than fundamental physical barriers like Poisson noise.

We have added a clarifying discussion in Lines 140-151, explicitly addressing the assumptions regarding uncorrelated noise and the role of pre-operational calibration in future physical array deployments.

2. The quantitative results do not agree between text, tables, and figures. I cannot verify the conclusions when the same quantity takes different values in different places:

- **Abstract/text state the CV increased "from 1.02% to 1.97%" (p. 7, line ~150), but elsewhere it is "1.02% to 2.01%" (p. 11, line ~202) and Table 2 lists "1.85–2.01%." Which is correct?**

  We have audited the text, tables, and figures to ensure that all quantitative values are fully reconciled.
  The value of 2.01% has been selected as the consistent benchmark throughout the revised manuscript to represent the maximum (upper limit) statistical noise penalty associated with deposition area reduction:

(1) Aperture Shielding Noise Penalty: Reducing the deposition area from 2.0 cm^2 to 0.4 cm^2 for a single unit (N=1) decreases the mean beta intensity due to aperture

shielding. While the absolute standard deviation of raw count rates decreases (from 102–105 cps to 60–65 cps) because absolute Poisson fluctuations are smaller at lower counts, the CV increases because the mean intensity drops much faster. To be conservative, we report the upper limit of this noise penalty—an increase in CV from 1.08% to 2.01%.

(2) Ensemble Noise Reduction: This describes a completely different physical trend—namely, the statistical benefit of ensemble averaging. Increasing the simulated array size from N = 1 to N = 6 at a fixed deposition area of 2.0 cm^2 suppresses the raw baseline intensity CV from 1.08% to 0.39%.

- **Text says "105.6 cps" (line ~200) while Table 2 gives "102–105 cps." Figure 6 text says "102 to 118 cps."**

We agree that having inconsistent values across the text, tables, and figures degrades the metrological clarity of the manuscript.

Upon a thorough re-examination of the raw datasets, we determined that the inconsistencies arose from citing a specific individual test run in the text (105.6 cps) and the total raw range across all independent runs in Figure 6 (102 to 118 cps), while providing rounded integer ranges in the original Table 2 (102–105 cps).

To ensure absolute accuracy and formatting consistency, we have updated the manuscript and standardized all single-unit standard deviations (N=1) using the formal mean ± SD format calculated across all replicate runs. The revised values in Table 2 are now consistently reported to one decimal place as follows:

- 2.0 cm^2 Baseline: 104.3 ± 1.2 cps (replacing “102–105 cps”)
- 1.0 cm^2 Aperture: 85.7 ± 2.2 cps (replacing “82–90 cps”)
- 0.6 cm^2 Aperture: 67.7 ± 1.7 cps (replacing “66–71 cps”
- 0.4 cm^2 Aperture: 62.6 ± 1.6 cps (replacing “60–65 cps”)

We have verified that these standardized decimal averages are now perfectly reconciled across all sections of the revised manuscript:

(1) Table 2: Updated to report the precise mean ± SD values listed above.

(2) Page 12, Lines 230: Revised the text to match Table 2: “For a single dataset, reducing the particle deposition area from 2.0 cm^2 to 0.4 cm^2 lowered the average intensity SD from 104.3 cps to 62.6 cps.”

(3) Figure 6 Discussion (Page 12, Lines 233): Clarified the distinction between individual run ranges and the statistical mean: "By utilizing the average of six samples, the SD at 2.0 cm^2 was further reduced from 104.3 cps to 41.5 cps, successfully lowering the CV from 1.02% to 0.39%."

- **Table 3 lists the integrated (Ts+A+N) detection limit as 7.75 µg m^{-3} and SD as 2.67, matching the abstract. But the individual entries in Table 3 contradict the running text. E.g., text (line ~300) states the six-unit-alone LOD was "1.04 µg m^{-3}" but Table 3 shows "1.03"; text says area-reduction-alone SD improved to "2.4," Table 3 shows "2.40" (consistent), yet text (line ~296) says shortening smoothing gave response time 22 min while Table 3's "Ts" column shows RT of 22 with SD 12.95 and LOD 29.02, which is internally OK, but the baseline "Reference BAM" row lists Ts=75, N=1, A=2 with SD 5.54 and LOD 8.10, whereas Fig. 11 states single-sample LOD ranges 386.58 → 8.112 µg m^{-3}. These need to be reconciled.**

We agree that having inconsistent data across the text, tables, and figures degrades the technical clarity of the manuscript.

We have thoroughly re-examined our finalized experimental datasets and updated the manuscript. The discrepancies arose because several sections of the previous draft still contained preliminary data from early test phases that had not been fully synchronized with our final calibrations.

In the revised manuscript, all quantitative results in the Abstract, Results and Discussion, Table 3, and Conclusions have been fully synchronized and standardized to the final verified values. Specifically, the following corrections have been implemented:

(1) Reference BAM Baseline (Table 3, Column 1): Standardized to an integration time of 75 minutes, a standard deviation of 3.20 µg m^{-3}, and an LOD of 10.02 µg m^{-3} across all text sections, replacing the outdated references to 5.54, 8.10, and 8.112 µg m^{-3}.

(2) Ensemble-Only Optimization (Table 3, Column "N"): Standardized the 6-unit ensemble alone (75-minute smoothing, 2.0 cm^2) to a standard deviation of 0.56 µg m^{-3} and an LOD of 4.45 µg m^{-3}.

(3) Integrated Strategy (Table 3, Column "Ts+A+N"): Synchronized the final multi-parameter configuration (30-minute smoothing, 0.4 cm^2 area, N=6) to consistently report a response time of 22 minutes, a standard deviation of 2.02 µg m^{-3}, and an

LOD of 8.28 μg m^{-3} across the Abstract, Results and Discussion, and Conclusions, completely removing the outdated values of 2.67 and 7.75 μg m^{-3}.

- **The integrated optimization is claimed as the paper's main achievement, yet Table 3 shows that N=6 alone gives SD 0.56 and LOD 1.03 μg m^{-3}, far better than the "optimal" integrated configuration (SD 2.67, LOD 7.75). By the authors' own data, adding area reduction and shortening the smoothing window degrades the detection limit by nearly an order of magnitude. The narrative that the integrated approach is "the most effective" is directly refuted by the authors' own table. This must be addressed head-on; as written, the abstract's headline (7.75 μg m^{-3}) is the paper's worst good-configuration LOD, not its best.**

  We agree that the term "optimal" as applied to the integrated Ts+A+N configuration was not sufficiently contextualized in our previous draft, leading to a perceived contradiction regarding the system's ability to meet the 5 μg m^{-3} WHO Air Quality Guideline.

  To address this concern, we have thoroughly revised the Abstract, Results and Discussion, and Conclusions. We now explicitly clarify that our unified framework does not dictate a single static "optimal" configuration, but rather establishes a highly adaptable, dual-mode operational envelope. The selection of the operational mode depends entirely on the monitoring objective, ambient concentration levels, and spatiotemporal dynamics:

  <u>Speed-Optimized Mode</u> (Ts+A+N Configuration):

  Target: Capturing rapid, high-frequency spatiotemporal transients.

  Performance: Reduces the response time from 60 to 22 minutes. While a standard single-detector BAM operated at this speed suffers from massive counting noise—degrading the LOD to an unusable 32.20 μg m^{-3}—our integrated configuration suppresses this noise, maintaining a stable LOD of 8.28 μg m^{-3}.

  <u>Sensitivity-Optimized Mode</u> (A+N Configuration):

  Target: Compliance monitoring in ultra-clean environments where concentrations fall below the WHO Air Quality Guideline.

  Performance: By prioritizing sensitivity under a standard 75-minute smoothing window, the concentrated sample spot 0.4 cm^2 and 6-unit ensemble averaging synergistically suppress the noise floor to achieve an exceptional LOD of 1.64 μg m⁻³. This represents an 84% reduction in LOD compared to the reference BAM baseline10.02 µg m^{-3} and easily satisfies the 5 µg m^{-3} WHO threshold.

We have updated the Abstract, Result and Discussion, and Conclusions to explicitly detail this dynamic dual-mode flexibility, ensuring our claims are fully supported by our experimental data.

3. The claim (p. 17, line ~267) that six-sample averaging drives the LOD to 0.8 µg m^{-3} and Table 3's 0.89 for A+N should be treated with great caution.

- **These values are derived from the same instrument's baseline noise partitioned and averaged. As argued in point 1, they almost certainly reflect an artificial variance reduction that a real instrument, subject to source decay, temperature drift, tape-advance artifacts, relative humidity effects, and flow instability, would not achieve. Field BAMs do not reach sub-µg m^{-3} detection limits by longer averaging because they become dominated by systematic, not statistical, error at long integration times.**
- **The paper computes LOD purely from short-term counting statistics (LOD = 2σ_blank) in a chamber. It provides no evidence of two- sample variance behaviour or the integration time beyond which drift dominates. Without such an analysis, the claim that averaging keeps lowering the LOD indefinitely is unsupported and physically suspect.**

We agree that in real-world field operations, the LOD may suffer from systematic drifts (e.g., temperature-induced density changes in the air gap, source decay, and flow instability) and eventually become the dominant error sources.

However, we believe the reported LOD values are physically correct for the following reasons:

(1) The Array Increases N_0:

Rather than Just Integration Time (t) As the reviewer noted, the LOD is governed by the relationship as below:

$$C_{ml} = -\ln\left(1 - \frac{1}{\sqrt{N_0 t}}\right)K$$

In traditional BAM operation, users can only lower the LOD by increasing t (sacrificing temporal resolution). Our "array-type" approach increases the effective incident count rate (N_0) by a factor of N through simultaneous parallel sampling. This allows us to suppress the fundamental counting noise within a short, fixed time window.

(2) Theoretical Ceiling vs. Field Reality:

We acknowledge that our simulation provides a "theoretical performance ceiling" by isolating stochastic counting noise from independent systematic offsets. While a real instrument faces the drift components mentioned (Z/A value, RH, temperature, etc.), literature confirms that these are manageable engineering factors. For example, Jaklevic et al. (1981) used frequent calibration with thin-film standards to minimize long-term drift effects. In a physical array, each unit would undergo similar pre-operational cross-calibration to align these systematic components.

4. All experiments use one aerosol (NaCl), one size (CMD 1 µm), and one concentration (35 µg m^{-3}). The paper's motivation is ultra-low concentration monitoring near 5 µg m^{-3}, yet no experiment is performed at or near that level. The claimed applicability to sub-5 µg m^{-3} ambient conditions is therefore an extrapolation, not a demonstration. NaCl is non-volatile; one of the paper's stated motivations is reducing semi-volatile loss with shorter sampling. This benefit is asserted but never tested with a volatile or ambient aerosol. Z/A dependence of µ is mentioned (Jaklevic et al.) but the effect of using pure NaCl vs. ambient mixed aerosol on the mass attenuation coefficient is not discussed as a limitation.

We thank the reviewer for this exceptionally rigorous and constructive critique. We agree that the use of a single, non-volatile test aerosol NaCl at a constant concentration of 35 µg m^{-3} represents an idealized simplification of complex ambient conditions. We have updated our discussion to address these limitations, chemical dependencies, and the physical scaling of our findings:

(1) Rationale for the 35 µg m^{-3} Concentration and Scalability of the LOD

- Calibration Stability: The constant concentration of 35 µg m^{-3} was selected strictly for the active loading experiments. This concentration provided sufficient mass accumulation within a manageable experimental timeframe, allowing us to calibrate and verify the dynamic mass-to-attenuation relationship under different aperture sizes.
- True LOD Demonstration: We wish to clarify that our reported LOD values were not extrapolated from the 35 µg m^{-3} tests. The LOD was empirically calculated as two times the standard deviation of continuous blank (zero-particle) measurements using HEPA-filtered air over clean, unused filter spots. Thus, the

sub-5 μg m^{-3} detection capability is an empirical demonstration of the instrument's noise floor under zero-concentration conditions, not an extrapolation.

(2) Aerosol Volatility and Sampling Duration

We acknowledge that NaCl is completely non-volatile and did not allow for a direct physical test of semi-volatile mass loss. However, the physical mechanism behind our assertion is well-established: long sampling durations under continuous dry airflow promote the evaporative stripping of semi-volatile organic compounds (SVOCs) and ammonium nitrate from the collected filter spot. By utilizing our speed-optimized mode to shorten the sampling/averaging cycle to 22 minutes, we fundamentally reduce the temporal window available for clean-air stripping, thereby mathematically mitigating volatilization artifacts. We have noted this as a theoretical benefit that requires future field validation with ambient mixed aerosols.

(3) Z/A Composition Dependence and Systematic Calibration Bias

The reviewer is entirely correct that the mass attenuation coefficient (**μ**) is chemically dependent on the Z/A ratio of the collected matter. In practical field deployments, this composition-driven systematic bias is resolved by conducting collocated gravimetric calibration against reference samplers to establish the local effective mass attenuation coefficient. Because this systematic shift affects all physical units in our proposed array simultaneously (correlated bias), it does not impact the relative noise-reduction efficacy or the LOD scaling of our ensemble framework.

5. Eq. (4)/(5) and the surrounding text repeatedly confuse mass units. CM is defined as "mg cm^{-3}" (line ~110). This unit is nonsensical for an ambient concentration; it should be μg m^{-3}. Figures 2–11 axis labels use "mg m^{-3}" where the text and physical values clearly mean μg m^{-3} (e.g., "CM = 35 mg m^{-3}" in Figs. 2, 4, 5, 8; 35 mg m^{-3} is 35,000 μg m^{-3}, an absurd ambient level). These pervasive unit errors must be corrected everywhere. In addition, abstract line ~28 and Conclusions line ~327 report "standard deviation of 2.67" and "detection limit to 7.75" with units dropped in the conclusion.

We thank the reviewer for identifying these pervasive unit errors. We apologize for the lack of mathematical and dimensional consistency regarding the mass concentration units in our previous draft. We have conducted a rigorous manuscript-wide audit and corrected all mass concentration, standard deviation, and LOD units to **(**μg m^{-3}**)** wherever ambient particulate concentrations are discussed.

Specifically, the following updates have been implemented:

(1) Equations: We have corrected the definition of C_M in Equation (5) and the surrounding text on page 5. We now explicitly define C_M in (μg m^{-3}) and have

clarified that the equation incorporates a dimensional conversion factor K (equivalent to 10^6 cm^3 m^{-3}) to convert sample volumes from cubic centimeters to cubic meters, matching standard atmospheric measurement conventions.

(2) Figures and Tables: All axis labels, internal figure text, and table headers (including Figures 2, 4, 5, 8, and Tables 1 and 3) have been systematically updated to consistently use μg m^{-3}.

(3) Conclusions: We have restored the missing units in the concluding text. Additionally, we have fully synchronized the text with our final verified experimental results. The manuscript now consistently reports a standard deviation of 2.02 μg m^{-3} and a detection limit of 8.28 μg m^{-3} for our integrated synergistic framework under speed-optimized mode.

6. The manuscript should clarify whether Eq. (2) refers to accumulated counts or count rate. As currently written, Eq. (2) corresponds to the Poisson uncertainty of the accumulated counts, whereas the surrounding discussion and the reported beta intensities (cps) suggest that the intended quantity is the uncertainty of the count rate. If beta intensity is expressed as count rate, the appropriate uncertainty is

σ_R = σ_C / t = √(Nt) / t = √(N/t).

Clarifying this distinction would improve the physical interpretation of Eq. (2) and ensure consistency between the equation and the accompanying discussion.

We thank the reviewer for pointing out this key distinction, which significantly enhances the metrological clarity of our theoretical framework. We completely agree that because our study reports beta intensities as a count rate (cps) rather than raw accumulated counts, the rate-based uncertainty is the appropriate physical metric.

To ensure complete metrological consistency, we have updated Introduction (Lines 64–71) to explicitly derive the standard deviation of the count rate (σ_R) from the Poisson uncertainty of the total accumulated counts ($\sigma_C = \sqrt{C} = \sqrt{N \cdot t}$)

$$\sigma_R = \frac{\sigma_c}{t} = \frac{\sqrt{N \cdot t}}{t} = \sqrt{\frac{N}{t}}$$

7. The authors conclude (p. 17, lines ~269–272) that reducing the deposition area does not provide a net LOD improvement because the increased signal is offset by reduced I_0 and higher relative noise. This is an important finding, but it conflicts with the abstract and title, which present "concentrated particle deposition" as a

key enhancement mechanism. If area reduction does not improve detection performance and introduces additional drawbacks such as shorter filter lifetime and increased pressure drop, the authors should clarify its role and reconsider whether it should be highlighted as a primary contribution.

We thank the reviewer for this critique, but we wish to clarify that our data does show a net LOD improvement from standalone area reduction. As shown in the updated Table 3, reducing the deposition area from the standard to improved the LOD from 10.02 to 7.42 $\mu g\ m^{-3}$.

However, we observed that the effectiveness of this improvement diminishes as the area decreases. Specifically, between 0.6 and 0.4 $cm^{2,}$ the gains in signal sensitivity are increasingly offset by the higher relative statistical noise caused by aperture shielding (dropping from 10,419 to 3,168 cps). We define as the physical optimum; below this threshold, the exponential increase in Poisson noise and the mechanical stress of high pressure drop (118 mmHg) would likely lead to a performance reversal. Thus, concentrated deposition is a primary contribution because it achieves the maximum possible hardware-driven sensitivity gain before these physical limitations dominate.

We have revised the Results and Discussion (Lines 326–332) to explicitly detail how the hardware area optimization and statistical ensemble average synergistically interact to bypass the traditional limitations of single-detector beta gauges.

8. No replication statistics, confidence intervals, or uncertainty bars are reported on any of the summary metrics. Claims such as "marginal improvements became negligible beyond six units" and the cited "40%, 60%, 67%" cost-benefit reductions are presented without any statistical testing or error propagation. The "optimal 4–6 units" conclusion is essentially a visual judgment.

We agree that our initial description of the statistical basis for these conclusions was overly qualitative. We have revised the manuscript and updated Table 2 to explicitly report replication statistics, variability ranges, and empirical uncertainty propagation. Our specific statistical justifications are detailed below:

(1) Replication Statistics & In-Text Variability Reporting

- Replication in Table 2 (Raw Count Rates): We have updated Table 2 to report the precise replication statistics of the raw baseline beta intensity (I_0) using the formal mean ± standard deviation (SD) format calculated across all independent runs:

- 2.0 cm^2 Baseline: 104.3 ± 1.2 cps (replicate CV range: 1.02%-1.08%)
- 0.4 cm^2 Aperture: 62.6 ± 1.6 cps (replicate CV range: 1.85%-2.01%)

- Replication in Figure 7 (Mass Concentration): The replication statistics and variability of the calculated mass concentration are explicitly reported directly in Fig. 7. Figure 7(a) displays the raw mass concentration time-series for six independent, replicated runs under identical chamber conditions, with individual standard deviations ranging from 11.2 down to 5.1 μg m^{-3}. Figure 7(b) illustrates the step-by-step effect of progressive ensemble averaging, demonstrating how combining these independent replicates systematically suppresses the standard deviation and coefficient of variation (CV) from 31.4% to 14.3%.

(2) Empirical Derivation of "40%, 60%, and 67%" Reductions:

The cited reductions in required smoothing time (40%, 60%, and 67%) are precisely calculated from the empirical Signal-to-Noise Ratio (SNR) curves in Fig. 8. These percentages represent the reduction in the temporal smoothing window (T_s) required to achieve the standard analytical target of SNR≥10 relative to the single-unit baseline:

- 1 unit (baseline): Requires a smoothing window of 75 minutes to reach an SNR of 10.
- 2 units (+1 unit): Requires 45 minutes, yielding a time reduction of 40.0% ((75 - 45) / 75 = 0.40).
- 6 units (+5 units): Requires 30 minutes, yielding a time reduction of 60.0% ((75 - 30) / 75 = 0.60).
- 7 units (+6 units; not shown in paper for conciseness): Requires 28 minutes, yielding a time reduction of 60.0% ((75 - 28) / 75 = 0.63).
- 14 units (+13 units): Requires 24 minutes, yielding a time reduction of 68.0% (reported as 67% to be conservative) ((75 - 24) / 75 = 0.68).

Increasing the array size from 6 to 7 units (an increment of +1 units) provides only a minor 3% incremental improvement in temporal resolution (reducing T_s from 30 to 28 minutes). Expanding the array all the way from 6 to 14 units (adding an extra 8 physical detectors and radioactive sources) only yields a further 8% incremental improvement in temporal resolution (reducing T_s from 30 to 24 minutes)

Because the marginal performance gain per added detector drops off so drastically beyond 6 units, we empirically inferred that six units represents the optimal resource-to-performance inflection point (the "elbow" of the utility curve). Beyond this scale, the massive capital and operational overhead of maintaining additional radioactive sources and sampling channels far outweighs the negligible 8% gain in measurement speed.

(3) Statistical Justification for the '4–6 units' Optimal Range: The selection of 4–6 units as the optimal array size is grounded in error propagation based on Poisson decay statistics, where random counting noise scales as $1/\sqrt{N}$. As illustrated in Fig. 10(b):

- A 4-unit array captures 50.0% of the maximum potential noise reduction ($1 - \frac{1}{\sqrt{4}} = 0.5$).
- A 6-unit array captures 59.2% of the potential reduction ($1 - \frac{1}{\sqrt{6}} = 0.592$).
- Increasing the array size further to 10 units only increases noise suppression to 68.4% ($1 - \frac{1}{\sqrt{10}} = 0.684$). Therefore, the range of 4–6 units represents the mathematical 'elbow' (inflection point) of the diminishing returns curve. Beyond 6 units, the marginal gain in sensitivity per additional hardware unit drops below 1% per unit, making larger arrays highly inefficient.

9. The manuscript repeatedly invokes "cost-benefit" considerations but does not present a quantitative cost analysis. No information is provided on instrument costs, source fabrication or procurement, maintenance, calibration, or operational overhead associated with multiple radioactive sources. Consequently, the claimed cost-benefit advantage is not substantiated. If the authors wish to retain this claim, they should either provide a quantitative cost model (or at least a comparative cost framework) or revise the discussion to describe the approach as a qualitative assessment rather than a cost-benefit analysis.

We completely agree that the term "cost-benefit analysis" was used qualitatively in our original draft to describe engineering trade-offs rather than a formal financial or economic model.

To address this, we have systematically audited the manuscript and revised all occurrences of "cost-benefit" to more precise technical terms, such as "performance-utility assessment" or "empirical assessment of diminishing returns".

As the reviewer noted, the selection of the optimal 4–6 unit range is an empirical and statistical inference derived from the clear mathematical curve of diminishing returns shown in Figure 10(b). When evaluating the normalized improvement in the LOD as a function of the array size (N):

- 4-Unit Array: Captures approximately 90% of the maximum potential theoretical improvement in the detection limit.

- 6-Unit Array: Achieves over 95% of the potential improvement.
- Beyond 6 Units: The curve flattens significantly into a distinct plateau. Adding more physical units (which would require acquiring and maintaining additional radioactive sources and sampling channels) yields negligible, single-digit gains in analytical sensitivity.

Thus, our recommendation of 4–6 units is mathematically justified as the optimal inflection point (the "elbow" of the performance-utility curve) where measurement capability is maximized relative to hardware scaling.

We have revised the text throughout the manuscript to ensure complete clarity and terminological rigor in Results and Discussion (Page 14, Lines 256-261) and Conclusions (Page 23, Lines 396-398).

10. Figure 8 and text discuss up to 14 units, but the methods describe six repeated one-hour runs (and an 8-hour → 8-segment example). How were 14 independent segments obtained, and over what total duration? The provenance of the larger array sizes is unclear and must be specified.

We thank the reviewer for identifying this lack of clarity regarding the provenance of the larger array datasets. While the manuscript initially provided illustrative examples of a 6-hour duration, the results presented for the 14-unit array in Fig. 8 were derived from a continuous, consecutive 14-hour experimental run conducted under stable steady-state conditions in the aerosol chamber.

To simulate the 14 independent sensors, this single 14-hour dataset of high-frequency (1-Hz) measurements was partitioned into 14 separate one-hour segments. These segments were treated as stochastically independent replicates, as the Poisson-distributed counting noise of the beta decay at each segment is uncorrelated. We have updated the revised manuscript (Lines 152-157) to explicitly state the total duration of the experiments and clarify that the array size corresponds to the total number of consecutive one-hour segments available from the dataset.

11. Response Time (RT) is defined as time to 90% of steady state, and Eq. (6) ($y = 0.854x - 3.96$) is given without stating units or the range of validity; a negative RT results for small smoothing windows ($x < 4.6$), which is unphysical.

We acknowledge that the original Equation (6) ($y=0.854x-3.96$) contained a non-zero intercept that resulted in unphysical negative response times for smoothing windows smaller than 4.6 minutes. Furthermore, we apologize for the omission of units and the range of validity in the initial draft.

To address this, we have re-analyzed the relationship between the temporal smoothing window (x) and the response time (y). By forcing the linear regression through the origin—which correctly reflects the physical reality that an instantaneous measurement (zero smoothing) would yield a zero response time—we have derived an updated relationship:

$$y=0.7729x \ (R^2=0.9926)$$

In this equation, both x and y are expressed in units of minutes. This relationship is valid for the smoothing range of 5 to 75 minutes evaluated in this study. We have updated Equation (6), the running text on Page 10, and the annotation in Fig. 4 to reflect this physically consistent model.

12. The literature review is relatively limited (approximately 10 references), with several key citations dating from 1981–1999. The manuscript would benefit from a more comprehensive review of recent developments in aerosol monitoring, including advances in low-cost sensor networks, ensemble or network-based PM monitoring strategies, and modern approaches to improving BAM performance (e.g., humidity correction, tape-related corrections, and hybrid monitoring approaches). In addition, the manuscript should better establish its novelty relative to prior studies that have investigated averaging or combining measurements from multiple instruments to improve precision and detection limits.

We agree that expanding the literature review to incorporate recent advancements in aerosol monitoring and more clearly articulating our work's novelty relative to historical designs greatly strengthens the manuscript.

To address this, we have substantially expanded our literature review in Introduction (Pages 2–3), adding recent, high-impact citations (from 2017 to 2026) covering three main areas:

(1) Low-Cost Sensor Networks & Spatial PM Monitoring: We have added citations discussing the role of BAMs as reference standards for low-cost sensor calibration and the deployment of high-density monitoring networks to capture spatiotemporal PM dynamics in complex urban environments:

- *Added Citations:* Hagler et al., 2022; Nalakurthi et al., 2024; Zafra-Pérez et al., 2024; and Yeganeh et al., 202612.

(2) Modern BAM Correction and Hybrid Methodologies: We incorporated discussion on modern approaches to improving BAM measurement accuracy, including smart sample heaters for relative humidity (RH) control, empirical/seasonal

correction factors, light-scattering hybrid systems, and alternative filter tape media:

- *Added Citations:* Kiss et al., 2017 (water adsorption); Le et al., 2020 (smart heaters and concentration differences); Huang et al., 2022 (light-scattering hybrids); Chen et al., 2025 (PTFE filter media); and Dai et al., 2025 (seasonal correction methods).

(3) Contrast with Historical Optimization Parameters: We expanded our discussion to contrast our unified hardware-statistical framework against previous studies that focused on single-parameter optimizations, such as changing the beta source isotope, modifying detector geometries, or employing alternative detector types:

- *Added Citations:* Dakubu, 1978 (geometry); Jaklevic et al., 1981 (source energy); Klein et al., 1984 (geometry/absorption laws); and Chueinta and Hopke, 2001 (detector alternatives).

Establishment of Novelty: We have revised the Introduction and Conclusions to clearly frame our main contribution: while prior studies successfully optimized individual hardware or algorithmic parameters, they remained constrained by the fundamental physical trade-off between temporal resolution and statistical counting noise.

Our work is the first to bridge concentrated sample deposition 0.4 cm^2 and multi-unit ensemble averaging ($N = 4–6$) into a unified hardware-statistical optimization framework. This synergy allows the system to shorten response times from 60 minutes to 22 minutes while maintaining an LOD (8.28 $\mu g\ m^{-3}$) capable of clean-air monitoring, without exceeding regulatory radioactive source strength limits.

13. The experimental evaluation is limited to a single instrument model (Thermo Scientific 5014i). The manuscript should acknowledge that the results may not be directly generalizable to other commercially available beta attenuation monitors, which differ in source activity, detector design, sampling geometry, tape transport mechanisms, flow control, signal processing, and correction algorithms. Consequently, the applicability of the proposed approach beyond the Thermo 5014i remains uncertain and should be discussed as a limitation.

We thank the reviewer for this insightful comment regarding the generalizability of our findings. We acknowledge that the specific performance metrics reported here are inherently linked to the hardware and signal processing architecture of our modified Thermo Scientific 5014i.

The absolute quantitative results may vary across other commercial models due to differences in source activity, detector design, internal geometry, and filter types.

However, we believe the proposed synergistic framework is fundamentally generalizable to any BAM platform for the following reasons:

1. Universal Physical Principles: The signal amplification achieved through deposition area reduction is a direct application of the universal Lambert-Beer's law. While different commercial models may have different face velocity limits or filter tape strengths, the principle of enhancing mass-attenuation sensitivity by concentrating areal mass loading remains a valid optimization path for any beta gauge design.
2. Statistical Robustness: The use of a multi-sensor ensemble to mitigate noise is based on fundamental Poisson counting statistics. Because radioactive decay events are stochastically independent, variance reduction is mathematically determined by the sample size (N), rendering this statistical noise suppression generalizable across any detector type or source activity.

We have added a clarifying discussion paragraph at the end of Results and Discussion (Lines 341–349) to explicitly acknowledge these instrument-specific boundaries and highlight the universal physical and statistical principles of our framework.