

General comments:

This manuscript presents a substantial field evaluation of Sensirion SPS30 and SEN55 PM2.5 sensors under outdoor and indoor conditions in Arba Minch, Ethiopia. The dataset is valuable, particularly because it includes multiple units of both sensor types, gravimetric comparisons, assessment and measurements in high-concentration biomass-burning conditions, and an analysis of the SEN55 digital truncation behavior. The comparison with the Hyltemossa dataset also provides useful evidence that some of the observed inter-model differences are not unique to the Ethiopian campaign and hence would be useful to the broader scientific community interested in using these lower-cost sensors.

Overall, I find the study potentially suitable for publication in *Atmospheric Measurement Techniques*. My main concerns are not with the presented experimental dataset itself, but with the strength and scope of some of the claims concluded from it. In particular, the manuscript sometimes moves from demonstrating good empirical agreement under the conditions studied to broader claims of validation, reference-equivalent performance, and accuracy at concentrations that were not independently resolved by the reference measurements.

I therefore recommend **major revision**, mainly involving further clarification and sentence reframing rather than doing any new experiments.

Major comments:

1. I recommend distinguishing calibration performance from independent validation

The manuscript applies sensor-specific bias corrections and then reports substantially improved accuracy metrics. This is useful and demonstrates the potential benefit of local calibration. However, it is important to distinguish performance after calibration from independent predictive validation. Please clarify which observations were used to derive each correction factor, which observations were subsequently used to evaluate the corrected data, whether the corrections were sensor-specific or pooled together, and whether any independent temporal, event-level, or location-level validation was performed.

If the same observations were used both to derive and evaluate the correction, the resulting post-correction accuracy should not be presented as an independent validation estimate. In that case, I suggest consistently describing it as performance after calibration or after correction. If feasible, a simple cross-validation analysis would strengthen the study, but I do not feel that a complete reanalysis is essential if the terminology and interpretation are corrected.

2. The conclusions concerning concentrations above $1000 \mu\text{g m}^{-3}$ should be narrowed

The analysis convincingly shows that truncating or removing measurements above $1000 \mu\text{g m}^{-3}$ substantially worsens agreement with the time-integrated gravimetric measurements. However, this does not independently establish the absolute accuracy of individual high-frequency sensor concentrations. The gravimetric measurements are time-integrated, and the SPS30 used to corroborate SEN55 truncation is itself an optical sensor.

I therefore suggest moderating statements such as “true physical occurrences” and “validated for the full range of concentration levels.” The results support the conclusion that retaining high-concentration measurements improves representation of integrated PM_{2.5} exposure, while the accuracy of the individual extreme peaks remains unverified.

3. The novelty statement should be reconsidered

The manuscript currently states that previous SEN55/SEN54 studies did not compare SEN55 with reference-equivalent instrumentation. This statement appears too broad and should be revised, particularly in light of several relevant studies (mentioned below)

“Field testing of low-cost particulate matter sensors for Digital Twin applications in nanomanufacturing processes”

“Monitoring and Ensuring Worker Health in Controlled Environments Using Economical Particle Sensors”

“Calibration of SPS30 Particulate Matter Sensor Using Feedforward Neural Networks” (For SPS 30)

The present study has clear value in its Ethiopian field setting, biomass-burning conditions, very high PM_{2.5} concentrations, and analysis of SEN55–SPS30 differences and digital truncation. I suggest reframing the novelty around these aspects rather than presenting the reference comparison itself as an unaddressed research gap.

4. Clarify the statistical independence of the sensor–gravimetry comparisons

The manuscript should clarify the statistical unit used in the regression and significance analyses. Multiple sensors appear to have been measured during the same gravimetric sampling periods, so the reported sensor–gravimetry pairings may not represent fully independent environmental observations, particularly for the pooled “All SPS30” and “All SEN55” analyses. Please clarify how this dependence was handled when calculating confidence intervals and testing differences between sensor types. If the observations were treated as independent, the potential clustering by sampling period should at least be acknowledged for the readers.

Minor comments

5. Please use “reference-equivalent” cautiously. Meeting a specified performance criterion should not necessarily be interpreted as establishing formal equivalence to a reference method; I suggest wording this as performance “within the relevant criterion.

6. Clarify the interpretation of the indoor sensor-to-sensor differences

The indoor slopes show appreciable unit-to-unit variation, particularly among the SEN55 units. The manuscript provides an interesting observation that sensors located on the same box have very similar slopes, suggesting that local positioning and the highly heterogeneous concentration field may contribute to the observed differences.

I suggest emphasizing this point slightly more in the discussion. It is potentially useful practical information, as under strong local sources, deployment configuration may contribute to sensor-to-sensor variability.

7. Clarify the RH analysis and sensor self-heating

The manuscript provides a useful analysis of the SEN55 temperature/RH offset and its implications for the hygroscopic-growth correction. I suggest making the distinction between ambient/reference RH and the RH measured inside the SEN55 more explicit in the main discussion, since this is important for interpreting the RH analysis. The results appear sufficient to support a cautious statement that no clear RH dependence was observed under the conditions studied, rather than a broader conclusion that RH corrections are unnecessary.

8. Additional context on the high-concentration measurement conditions should be provided

Because the high-concentration indoor measurements are an important part of the study, please provide some additional information on the experimental setup, particularly the type of solid fuel used and the approximate height and location of the sensor inlet relative to the cooking stove. A brief description of the ventilation/cooking setup, where available, would also help readers interpret the extreme concentrations.