The revised manuscript by Bau et al. deals with an interlaboratory comparison of optical particle size spectrometers. For this study, an impressive number of partners and tools were made available The information provided makes the change to the heading worthwhile. Overall, the manuscript is well written, although there are still a few questions that need answering.

ABSTRACT

L33 "… instead of establishing or developing a calibration procedure…" - Am I right in thinking that this means a method is possible even without calibration, one that allows for comparability? Or that calibration isn't necessary? It could be interpreted either way.

> The end of the sentence dealing with the establishment of a calibration procedure has been removed in the revised version of the manuscript. Indeed, this approach requires the use of reference calibrated instruments and model particles which is clearly beyond the scope of this paper.

MAIN

L101 - Is this a sort of reliability test, then? Would a service or a calibration be in order now?

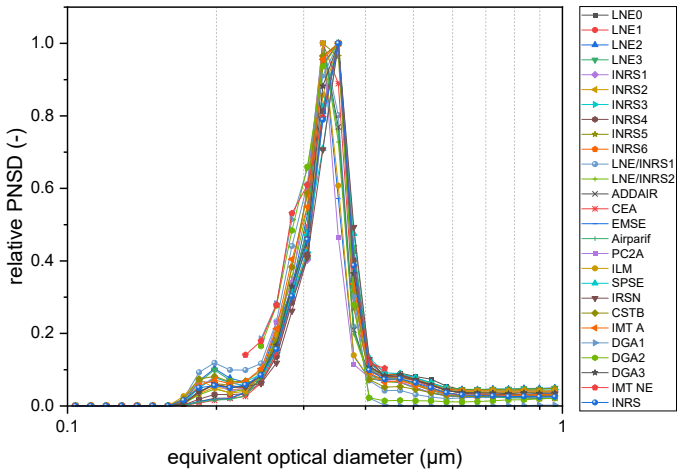
> Our data cannot be considered as a calibration, it consists of a metrological control as stated line 101. This inter-laboratory comparison provides a kind of inter-model variability and inter-specimen reproducibility.

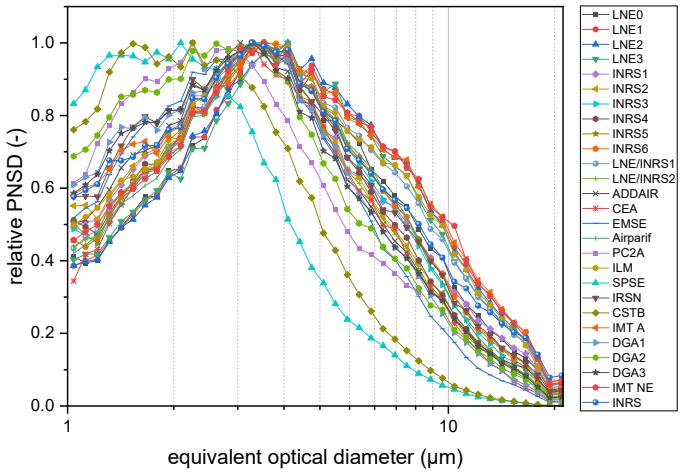
L113 - Can you explain to what extent this constitutes a reference dataset, given that no calibration data for the instruments was available (factors for ambient refractive index and more)?

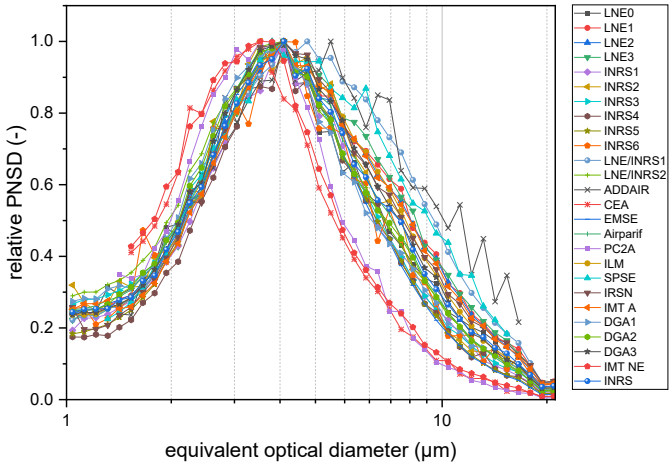
> The term "reference" has been changed into "first".

L120f - You can generate a summary OPSS number-size graph for each partner measurement, showing the OPSS comparison unit (that one that stayed always the same device) measurement for each comparison at that location, so that you can assess any fluctuations between measurements.

> According to your comment, please find below the graphs for samples A, B and C, for control OPSS.







The variability observed for the control instrument for both the modal diameter and the GSD obtained from lognormal fit are gathered in Table 2.

L168 “monodisperse” -In principle, yes, but in the later graphs larger particles can also be seen, which are likely agglomerates; in terms of total number, they probably account for as much as the main mode. Please also note that the atomisation method likely produces many agglomerates

Fig 3 shows that the monodisperse population of the silica sample (~ 0.3-0.4 µm in diameter) is largely the main population, though few agglomerates are also seen in the micron size range (1-10 µm). The atomisation mentioned in your question process has not been involved; only a dry-based method was used to produce the test aerosols as stated in section 2.1.

L169 “optical properties…” – which one? Refractive index?

According to your comment, refractive index is clearly mentioned in the revised version of the paper.

L187f “good laboratory practice” - Please explain what this means in this context. (Perhaps the length of the tubes, no right angles, etc.?)

What is meant is exactly what you suggest in your comment; it has been specified in the revised version of the manuscript.

L212 – “total number concentration” - Some of the instruments exhibit higher error rates the closer they get to their maximum count. Depending on the algorithm, only the large particles may then be counted, or larger bins may be counted even though none exist. In addition, miscounts can occur at low bins if very large particles (roughly three times the wavelength) are present in the system. Has this been accounted for in the evaluations?

Except lognormal fit, no specific processing was performed on data provided by the partners, especially dealing with number concentration. As stated a couple of times in the paper, “each participant was asked to make sure that each OPSS involved was used under “good laboratory practices” conditions, i.e. by ensuring the absence of internal errors reported by the devices, including coincidence errors.” To our point of view, miscounts and more generally correction algorithms inherent to each device contribute to the inter-model and inter-specimen variability.

Figure 4 - Taking the Grimm 1.108 as an example, the error appears to be highest here; however, the test aerosol is also relatively small and at the lower limit of the smallest size channel, which should be noted

Error bars presented in Figs 4-6 correspond to the 95% confidence interval on modal diameters and GSD stemming from the fitting procedure. As stated in section 2.2, sample A consists of calibrated monodisperse particles in the low particle size range of OPSSs. For the two first datapoints obtained for Grimm 1.108, the 95%-confidence interval on both the modal diameter and the GSD of the fitted distribution are actually larger than other OPSSs. This is due to increased uncertainty in the lognormal fitting protocol, though very satisfying determination coefficients ($R^2 > 0.9$).

Fig. 5 and 6 - Depending on the manufacturer, it is apparent that instrument types cluster closely in the comparison of GSD and mode; this could be due to the calibration that manufacturers use for the respective instrument; at Palas, the different devices are built and calibrated at different locations or procedures, or isn’t it?

We are not convinced that instrument groups can be pointed out from Figs 5-6. Instead, the results provided here show a typical inter-specimen variability. We have not investigated how instruments are calibrated by the different manufacturers. We have to remind here that all OPSSs were taken in their current state without intention to involve freshly calibrated instruments.

Fig.7 -Without or using a different control instrument, this would look very different indeed; what does that imply?

We would like to thank the Reviewer for raising this interesting point. The choice of the control instrument relied on (i) its availability, (ii) its size range relevant to PNSD to be measured and (iii) its high size resolution (64 channels), as stated in Table 1. This has been added in the revised version of the manuscript.

According to equation 1, Z-scores are particularly affected by the repeatability standard deviation. The ratios sigma/mean for modal diameters are 1.8%, 16.7% and 11.5% for samples A, B and C, respectively. A similar, but less pronounced, behaviour is observed for the GSD, with 1.0%, 5.7% and 6.9%, respectively. These coefficients of variations can be supposed equivalent if another control instrument had been chosen.

This is what is meant in our discussion lines 282-286: “This is due to a greater uncertainty stemming from the multiple measurements performed with the control OPSS, which lead to

lower Z values. On the contrary, the highly repeatable PNSD obtained for monodisperse silica results in larger Z-scores, with a large proportion (62% for the modal diameter, 42% for the GSD) in the questionable and unsatisfactory (|Z| > 2) categories linked to the size detection limit of each involved OPSS (180nm for PALAS and 250-300nm for GRIMM & TSI, see Table 1)."

CONCLUSION

L431- "need of strengthening robustness in…" Please clarify what you might still missing in regard of ISO standard 21501 that is meant for the standardization for OPSS measurements

Our results highlight several limitations that are not fully addressed within the current framework of the ISO 21501 standard. Although ISO 21501 provides essential requirements for the calibration and performance verification of optical particle counters under controlled laboratory conditions, it does not comprehensively address the influence of complex aerosol properties encountered during ambient or operational measurements, such as non-sphericity, particle agglomeration, variable refractive indices or broad polydisperse size distributions. In addition, the standard provides limited guidance regarding inter-instrument comparability under realistic field concentrations, long-term instrument drift, coincidence effects during high particle loading and traceability of calibration procedures, especially when alternative aerosol materials are used instead of reference PSL particles. The present intercomparison therefore emphasizes the need for complementary harmonized protocols capable of evaluating OPSSs performance under more representative atmospheric conditions, including standardized approaches for uncertainty estimation, maintenance practices and correction algorithms adapted to complex aerosols. To our point of view, such developments would strengthen the practical applicability of ISO 21501 standard and improve the comparability of OPSS measurements across laboratories and monitoring networks.

According to the Reviewer's comment, a sentence related to ISO standard has been added in the revised conclusion.

Please answer following questions:

How could this study be replicated without having the control instrument, since every measurement and provided data is relative to this one, or how does the dataset help if no absolute figures are given regarding number concentration density?

In our study, there was no intention to involve neither reference test aerosols (e.g. PSL), nor reference instruments (e.g. SMPS, APS), nor freshly calibrated devices, without target concentration levels. Regarding the latter point, the only constraint was participants to make sure they used their OPSSs in an acceptable concentration range, i.e. in "non-coincidence conditions". For that reason, no absolute figures are given, since particle concentration ranges very highly variable between participants.

This study aimed to promote straightforward approaches with versatile systems in order to quantify inter-model and inter-specimen variability, even in the absence of the control instrument and without information on aerosol concentration. The results provided in this paper shall be considered as a basis to further interpret possible deviations between OPSS, and provide any user an order of magnitude of possible biases. As stated in the conclusion, "this study offers an essential starting point for future work that will need to address the more advanced calibration approaches and to further investigate how such methods could be implemented in practice."