We would like to thank all Reviewers for their valuable feedback. Please find below the point-by-point response to the questions raised.

Referee Report 1

This study compares the performance of a range of particle counters for periodic technical inspection on different polydisperse aerosols, such as soot produced by combustion. With its methodology, content and conclusion it is an extension to a previous study done by Vasilatou et al (2023) "Effects of the test aerosol on the performance of periodic technical inspection particle counters". This review is written based on the track changed version/ reviewed first manuscript.

General Comments / Questions:

- Line 241: "… test aerosols such as NaCl…" – Figure 5 a and b show great deviation for all types of test aerosols and instruments. Figure 5a up to 50%. Nebulised NaCl tends to have a large variety of GMD (Global Mean Diameter), depending on its concentration of the nebulised Liquid (Liu and Lee, 1975). Since the counting efficiency of diffusion chargers depends on the size distribution (Fierz et al 2007), it is possible to have a fitting salt GMD as well. Vasilatou et al 2023 examined just one possible NaCl GMD.

Answer: We respectfully disagree with this comment. Vasilatou et al. (2023) examined four different salt generators (FCS 249, ATM 228, ATM 220 and Meinhard) and various GMD_{mob} in the range 30 - 120 nm (see Figs. 3, 5 and 7 of the respective publication). In the manuscript under review we focus, however, on test aerosols with a GDM of 80 nm as required by the Swiss, Dutch and Belgian legislation for the yearly PN-PTI verification (in Germany the requirement is GDM = 70 ± 20 nm).

- A Previous Reviewer (https://doi.org/10.5194/ar-2023-16-RC1) posted: "The recommendations given in Section 4 do not seem to be directly drawn based on the results of this study. Rather, some of the recommendations are generic and seem to be based on the results of other or previous studies" I would agree upon this statement! Please clarify this in the text, that the recommendations are done (Line 250) based on the results of this study and previous.

Answer: We have amended the sentence as follows: "Based on the results of this and previous studies (Vasilatou et al., 2023), the following recommendations can be made…".

- Recommendations key message:

-Recommendation 1: Vasilatou et al (2023): "…highly recommended to use particles for calibration purposes which are similar in chemistry and morphology to soot particles emitted by vehicle engines" and Hammer et al. (2023): PN-PTI counters should ideally be performed with soot as test aerosol"

Recommendation 2 is about the setup correction factors for both.
The Conclusions are short with limited discussions. Line 281: "This study confirms that soot aerosols …. Are more suitable than NaCl". The only data for this statement are taken from Vasilatou et al 2023! Even there the NaCl data is limited. No comparison for different NaCl GMD. And NaCl data is within the same range as Soot aerosol.

Answer: As mentioned previously, Vasilatou et al. examined four different nebulisers (FCS 249, ATM 228, ATM 220 and Meinhard) and various GMD_{mob} in the range 30 - 120 nm (see Figs. 3, 5 and 7 of the respective publication). We do not agree that the data are limited. The size distribution of NaCl aerosols **must** be within the same range as for the soot test aerosols. National legislations prescribe specific GMD_{mob} ranges for instrument type-examination and verification. Both studies were carried out in the particle size range defined by national regulations.

- Missed opportunity for a DMA -CPC to DMA - PN-PTI cut-off efficiency experiment. This study would benefit from this insight.

Answer: We thank the reviewer for this comment. Our scope was to examine the verification of PN-PTI instruments, which is exclusively carried out with polydisperse particles. Monodisperse soot aerosols are sometimes used for instrument type-approval, but this was out of the scope of the current study.

- Please add grid lines for all figures

Answer: Certain journals recommend not to use grid lines in the figures. If the journal editor advises is to add grid lines, we will be happy to do so.

Specific Comments:

- Line 17 – Abstract: … "0.25 units" … - do you mean "%" , in that case, Figure 3 and Figure 4 shows that this threshold was exceeded many times.

Answer: The ideal counting efficiency of an instrument is 1 (= 100%). 0.25 units refers to the counting efficiency of 1. This would be equal to 25 % if the counting efficiencies were expressed as 100%.

- Line 21 – Abstract: …"MISG may be satisfactory" – that conclusion is based on the instrument response and ET/TC mass fraction, but effective density and particle size are in clearly different. Please clarify.

Answer: This is clarified is Section 4: "Low-cost soot generators can be a stable source of combustion particles and can be employed for PN-PTI verification using the appropriate setup correction factors. However, the GMD they produce should be in the range 70±20 nm in order to comply with the current linearity verification requirements in Europe". Note that national legislations do not specify a range for the effective particle density of the test aerosols.

- Line 89 : "… 8 port flow splitter" – Are concentration adjustments based on line loss (towards the CPC) or flow/pathlength to each individual instrument be considered?

Answer: Thank you for bringing this up. We have added the following sentence to explain how we ensured that the losses are equal: "The length of the tubes from the flow splitter to the devices was adapted to the respective flow rate to ensure equal diffusion losses".

- Line 95 + 185: "… counting efficiency was taken into account" – As you mention it multiply times, I would assume, that you multiply the Particle Number Concentration with an correction factor, please clarify in the figures as well.

Answer: We would argue that the most suitable place to describe how the measurements were performed is in the Section "Materials and Methods" as we have already done.

- Line 204: "… duration of 2 min" – Why was 2 minutes chosen? Duration for a standard engine exhaust test?

Answer: Yes, this is explained in Lines 120-121: "… thus the duration was similar to the duration of real PN-PTI tests which varies from 15 to 90 s".

- Line 231: "… tends to be larger …"- nearly a factor of 2! Please clarify.

Answer: In Lines 265-267 we highlight the fact that "the properties of real diesel soot can also differ considerably, depending on the engine design, driving cycle and fuel properties (Hays et al., 2017; Wihersaari et al., 2020)". This is why PN-PTI instruments are type-examined in the particle size range 23 nm – 200 nm. Aerosols with a GMD_{mob} of 100 nm generated by the MISG fall within this size range and are very close to the recommended size range of 70 ± 20 nm for PN-PTI verifications.

- Line 256: "range of 70+-20 nm" you mention in the Abstract, that the MISG is suitable, but in your recommendations it wouldn't fit. Please Clarify

Answer: In the abstract, we state that "MISG **may** be a satisfactory - and affordable - option for PN-PTI verification". In Section 4 (Recommendations) we further explain that the MISG would need to fulfil the requirement 70 ± 20 nm to comply with the current linearity verification requirements in Europe. This might be possible in the near future by using a different fuel mixture that has not been tested so far or by stabilising the various flows to make generation of smaller particle size distributions more reproducible. To avoid any misunderstanding, we have amended the Abstract as follows: "… however, further optimization will be needed for low-cost soot generators to comply with European PN-PTI verification requirements".

References
Vasilatou, K., Wälchli, C., Auderset, K., Burtscher, H., Hammer, T., Giechaskiel, B. and Melas, A.: Effects of the test aerosol on the performance of periodic technical inspection particle counters, J. Aerosol Sci., 172(January), doi:10.1016/j.jaerosci.2023.106182, 2023.
Liu, B.Y.H., and K.W. Lee (1975) "An Aerosol Generator of High Stability" Am. Ind. Hyg. Assoc. J., 36, 861–865
Fierz, M., Burtscher, H., Steigmeier, P., and Kasper, M., "Field Measurement of Particle Size and Number Concentration with the Diffusion Size Classifier (Disc)," SAE Technical Paper 2008-01-1179, 2008, https://doi.org/10.4271/2008-01-1179.

Referee Report 2

General comments

This manuscript describes the relative response of PN-PTI instruments with a variety of soot sources including one Euro 5b diesel vehicle. Such a study is needed because non-soot sources (NaCl and spark-discharge) have been shown to produce instrument responses that greatly differ from soot. This study shows that a wide range of soot sources could be used to meet the specification requirements for PN-PTI calibration, and thus, this study should be published as it will help develop standards and legislation.

Specific Comments

- Line 85. Figure 1 shows a blower but it is not described in the text. Please state why it was used.

Answer: Thank you for pointing this out. We have amended the text as follows: " To deliver the aerosol into the mixing volume, a blower (Micronel AG, Switzerland) was used".

- Line 85. Figure shows that compressed air was added to the mixing chamber but it was not described in the text.

Answer: We have amended the text as follows**:** "...a custom-made dilution bridge, and was mixed and diluted with filtered air in a 27-ml-volume chamber".

- Line 85. It is a good idea to use a mixing volume as was done. However, verification that the aerosol is well mixed is after the flow splitter is also a good idea. Were checks made to ensure the aerosol was well mixed (i.e. rotating instruments between sampling ports)?

Answer: Thank you for bring this to our attention. We have added the following sentence: "The splitter bias was determined according to the procedure specified in the ISO 27891 standard and was found to be within 1 % for particles with a GMD_{mob} equal to or larger than 23 nm".

- Line 151; "The fractal dimension Df of soot particles with a nominal GMDmob of 100 nm was derived via image analysis ...." It's not clear how this was done. Was a DMA used to pre-classify the particles onto TEM grids, or did you assume that 100 nm projected area was approximately the equivalent to 100 mobility diameter? Please explain.

Answer: We have now repeated the analysis by size-selecting the soot particles with a DMA and Table 1 has been revised accordingly. The calculated Df of size-selected particles are more accurate than the values reported before, but since they still lie in the same range (1.55 – 1.65) the discussion does not change significantly.

We have also modified the text as follows:

Lines 128-129: "The fractal dimension D_f of size-selected soot particles with a mobility diameter d_p of 100 nm was derived via image analysis of high-quality TEM-images using the FracLac feature of ImageJ 1.53e (ImageJ, National institutes of Health, USA)."

Lines 186-188: The calculated fractal dimension of soot particles lied in the range 1.55 – 1.65 for all generators, in line with the fractal-like morphology observed in the TEM images and with previous studies on freshly emitted soot particles from different combustion sources (Pang et al., 2023).

In the case of the Euro 5b diesel vehicle, we could not repeat the measurement with size-selected particles, therefore we deleted the calculated Df value from Table 1.

- Section 3.1. This section is supposed to be for results, however, the authors put many experimental method details which are better placed in Section 2. For example, Lines 151 to 155.

Answer: The Reviewer is right. We have moved the text related to the determination of the effective density and fractal dimension into Section 2 "Materials and methods".

- Line 158. The effective density measurement method should be described in Section 2.

Answer: We have shifted the sentence: "The effective density was determined for the 100 nm setpoints using an Aerodynamic Aerosol Classifier (AAC, Cambustion, UK) and a DMA (TSI Inc., USA) in tandem as described in (Tavakoli and Olfert, 2014)" to Section 2.

- Line 160. The effective density results could be put into context by comparing them to summary work by Olfert and Rogak (https://doi.org/10.1080/02786826.2019.1577949). They show the 'average' effective of denuded soot is 0.51 g/cm^3 at 100 nm mobility diameter, although there is fair amount of spread in the data and CI engines tend to have higher effective densities.

Answer: Thank you for bringing this article to our attention. We have added the following text: "According to the summary work by Olfert and Rogak, the effective density of denuded soot from various sources (gas turbines, compression ignition engines and laboratory-based burners) lies typically in the range 0.4-0.8 g/cm^3 at 100 nm mobility diameter (Olfert and Rogak, 2019) Compression-ignition engines tend to produce soot with higher effective densities, while gas-turbine soot tends to have lower effective densities (Olfert and Rogak, 2019)".

Referee Report 3

General Comments

The paper describes a follow-up study of Vasilatou et al. 2023, focusing on the influence of various polydisperse soot aerosols for the in-field calibration of DC based PN counters for PTI of Diesel vehicles. It reveals that the aerosol properties affect the CE of the instruments under test. In the actual study, the authors focused on combustion generated soot particles. In the previous study, the influence on the CE was enormous when using various types of test aerosols (salt, soot-like). In this study, when using combustion generated soot particles only, the influence on the CE is still significant, but could not fully be explained by the particle morphology, that was claimed the main influence quantity besides the particle size. Based on the findings, the authors recommend the usage of one of the used soot generators for type approval and annual calibration.

Strengths of the paper

• This work revealed the importance of test aerosol properties, like the particle material, which was not addressed in the various device specifications for PN PTI measurement.

• The experimental methods were well-founded (except the choice of the reference instrument), and the properties of the test aerosols were well characterized.

Weaknesses of the paper

• The main flaw of the paper is, that it does not deliver an explanation of the influence mechanisms on the CE of the various instruments. There is absolutely no common trend towards the various instruments, therefore, it is not possible to draw general conclusions about DC based instruments or the usage of generators. The aerosol properties were investigated regarding the particle morphology quite detailed. At the beginning of Section 3.1 it is stated, that DC based instruments are generally affected by the particle morphology because the average number of charges within the diffusion charging process will vary with the particle shape. However, there is no deeper explanation on how the instruments' CE is affected by morphology. Other studies or explanation by the manufacturers should be considered for the paper!

Answer: We have discussed the results with various instrument manufacturers and they could not offer an explanation. In addition, the manufacturers have not published any information on the exact design of their sensors neither in peer-reviewed articles nor in the instruments' manual. Without knowledge of the sensor design, especially the type of charger and voltage applied, it is impossible to interpret the results. We would like to encourage the manufacturers to work together with the aerosol research community and discuss these issues openly, but we can also understand if they do not want to disclose any company secrets.

• The second point is, that the choice of the reference instrument is questionable, because it is just one PTI instrument and the bias of -23% against the METAS reference could be problematic for practical applications. There are at least two questions that need to be answered.

1. Was the reference instrument tested against other METAS references up to the required concentrations of 5E6 cm-3 using polydisperse particles?? (The ISO27891 does not include a linearity measurement with polydisperse aerosols!)

Answer: The NPET relies on a CPC sensor and is therefore a very reliable instrument, which yields reproducible measurements. As we have shown in a previous publication (https://doi.org/10.1016/j.jaerosci.2023.106182), the plateau efficiency of the NPET does not depend on the test aerosol material, making it a suitable reference instrument. Furthermore, the NPET was calibrated against the METAS primary standard for number concentration according to the ISO 27891 standard and its counting efficiency was taken into account during data evaluation. The linearity of the NPET was checked and was found to be very close to 1. We see absolutely no reason why the NPET cannot be used as a reference instrument for calibrating DC-based PTI counters in the field.

2. An explanation why the CE is up to -23% against the METAS reference must be given! Was there a traceable calibration by the manufacturer before? Why is there such a large difference? Was the instrument used before calibration and did it drift to such an extend?? If this instrument would be used as a reference in the field, a bias of -23% could be problematic since the maximum permissible error in some regulations (like the NPTI proposal) is 25% only!!! The "COMMISSION RECOMMENDATION of 20.3.2023 on particle number measurement for the periodic technical inspection of vehicles equipped with compression ignition engines" also states the requirement that the reference systems' MU shall be less than 20% for subsequent verification!

Answer: We thank the reviewer for this comment. NPET was chosen as reference instrument because in a previous study we found that its plateau efficiency did not depend on the test aerosol material (this information was added in the paper). Before the testing campaign, we calibrated the instrument. The values reported in the paper (deviations of 23 % etc.) were taken into account during the data evaluation and any bias related to NPET counting efficiency was eliminated. In order to avoid confusion, we now report in the manuscript the calibration factor we determined. In addition, the NPET linearity was almost 1 in the concentration range relevant to our study. Moreover, the commission recommendation refers to official PTI tests of diesel engines, not to research studies as the one under review. The goal of our study was to test and compare PTI instruments with different soot aerosols, not to type-approve diesel vehicles.

• There was no added value from the profound investigation of the test aerosol properties. None of the parameters listed in table 1 delivered an explanation of the CE behavior of the instruments under test.

Answer: The profound investigation of the test aerosol properties was considered necessary for explaining different CEs of DCs with different soot generators. Unfortunately, this was not possible due to the absence of information on the design of the counters and possible built-in corrections. However, we strongly believe that a profound investigation is useful for this research field as well as for the ongoing discussion on the design of DC sensors. The conclusions of this study may not give input on DC-sensor optimization but do point out possible sources of errors in PN-PTI and how they can be (at least) reduced.

• Is the test aerosol from one EURO5 vehicle representative for "Diesel soot" in general?

Answer: As already noted in the manuscript: "Further studies with more diesel test vehicles would be necessary to elucidate which type of laboratory-generated soot is the best proxy for diesel soot, keeping in mind that the properties of real diesel soot can also differ considerably, depending on the engine design, driving cycle and fuel properties (Hays et al., 2017; Wihersaari et al., 2020)".

- Specific comments - Line 93: reference instrument: Was this instrument tested against other METAS references up to the required concentrations of 5E6 cm-3?? (The ISO27891 does not include a linearity measurement with polydisperse aerosol)

Answer: Please see answer above related to the same question.

2. Please explain why CE is -23% against the METAS reference! Was there a traceable calibration by the manufacturer before? Why is there such a large difference? Was the instrument used before calibration and did it drift to such an extend?? If this instrument would be used as a reference in the field, a bias of -23% could be problematic, since the maximum permissible error in some regulations (like the NPTI proposal) is 25% only!!!

The " COMMISSION RECOMMENDATION of 20.3.2023 on particle number measurement for the periodic technical inspection of vehicles equipped with compression ignition engines" also states the requirement that the reference systems MU shall be less than 20% for subsequent verification!

Answer: Please see answer above related to a very similar question.

- Line 123: aerosol properties: The investigation of the aerosol properties was interesting and detailed but limited to the morphology of the particles. This is fine under the assumption that the particle morphology is the main parameter that affects the CE of DC based instruments. Are you sure that there is no other parameter, that affects the found behavior in the same order of magnitude?

Answer: We never claimed that particle morphology is the only parameter that affects the counting efficiency of PTI counters. On the opposite, we stated that: "Particle number concentration measured by diffusion chargers depends on the average number of charges carried by each particle (Fierz et al., 2011). Particle size and morphology have been shown to have an effect on the number of charges carried by the particles and, thus, on the counting efficiency of diffusion charger based PN-PTI instruments (see (Dhaniyala et al., 2011; Vasilatou et al., 2023) and references therein). Soot particles form complex structures described by a fractal-like scaling law (Mandelbrot, 1982), and their mobility is influenced by their morphology (described by the fractal dimension and fractal pre-factor) and the momentum-transfer regime (Filippov et al., 2000; Melas et al., 2014; Sorensen, 2011)". In this study, we performed experiments at different particle sizes and we characterised the test particles not only in terms of particle morphology, but also in terms of effective density, EC/TC ratio and fractal dimension.

- Line 189/190 & Line 195/196 It would be interesting to have an explanation how the CE varies with different GMDs and particle materials. To my impression, this very much depends on the measuring principle of each individual device, and thus, it would be good to explain the differences by the various measuring principles!

Answer: See answer to similar question above. The exact design and measurement principle of the sensors is not known.

- Line 202/203 This conclusion is questionable! An explanation of the scattering of the values must be delivered, rather than conducting the test at more comfortable test conditions!!

Answer: As already mentioned in the manuscript, the performance of DC-based sensors depends on many different parameters (e.g. particle size, morphology, particle charge etc.) and it is impossible to disentangle the effects of each aerosol property on the counting efficiency of the sensors, especially when information on the design of the instruments is not available. But even if one could explain why the PN-PTI instruments show different counting efficiencies, this would not change the fact that the values scatter more at particle sizes larger than 90 nm. Our conclusion that "This supports the choice of soot with 50-90 nm mobility diameter for the PN-PTI instruments verification linearity tests" is therefore valid.

- Figure 5b) The comparison with Vasilatou et al., 2023 shows, that there might be other influence quantities than the particle morphology only. The importance of pre-existing charges is obvious in figure 5b) and might be even more important if there are internal correction factors used, as in the case of CAP3070! I recommend to read Knoll et al., 2021 regarding influence of pre-charges on DC based instruments.

Answer: We fully agree that initial particle charge is important. This has been discussed in Vasilatou et al. (https://doi.org/10.1016/j.jaerosci.2023.106182), where the results shown in Fig. 5b are taken from.

- Line 297/298 The link to this reference does not work!

Answer: Thank you for bringing this to our attention. It seems that the page has been removed from the EU website. We have revised the reference as follows: Anon: Proposal Particulate Number Counters, [online] Available from: https://nmi.nl/special-particle-number-counters/, n.d.

- Line 204 / Figures S6 – S9 in supplement There were some significant fluctuations of the CE signal using the differnet aerosol generators:

1. The DX280 and the Knestel had big fluctualtions with the MiniCast 5201 under fuel rich conditions 2. The fluctualtions of all instruments, except the HEPaC and the CAP3070 were very large with the MiniCAST 6204 3. The AEM had fluctuations of about 50% CE using the MISG without cyclone --> Please explain why those signals were so unstable!!! --> What was the reason (pressure fluctuations, flow rate, heating, signal processing of instruments, etc.) and why is there no clear trend towards the instruments and the generators used?

Answer: The environmental conditions in the laboratory were very stable, so was the aerosol generation. The fluctuations are due to the PN-PTI instruments. Note that these are mid-cost instruments designed with the PTI test in mind, i.e. to check whether diesel vehicles emit above the limit value (e.g. 250'000 cm^{-3}). Considering that the reported number concentration is typically averaged over 30-90 sec, the observed fluctuations have little to no influence on the end result, i.e. whether a diesel vehicle will pass or fail the PTI test.

- Can you exclude instabilities of the reference instrument? Further literature to be considered - Bainschab et al., 2020 (https://doi.org/10.1016/j.aeaoa.2020.100095) Calculation of false pass / false fail scenarios and general impact of PN PTI on fleet emissions. - Krasa et al., 2023 https://tandf.figshare.com/articles/journal_contribution/Toward_a_simplified_calibration_method_for_23_nm_automotive_particle_counters_using_atomized_inorganic_salt_particles/22121581 For the sake of completeness in the introduction, the study of Krasa et al., 2021 should be mentioned. It shows a significant impact of the test aserosol also for CPCs. - Knoll et al., 2021 (https://doi.org/10.1080/02786826.2021.1873910) Influence of pre-charges on DC based instruments.

Answer: The NPET has been calibrated multiple times against our primary standard at METAS the past few years and has shown very good stability and reproducibility. Any instabilities of the NPET are negligible compared to the other measurement uncertainties.

Additional correction by the authors: We have spotted an error in Figure 5 and revised it so that the values are in agreement with the measurements published in Vasilatou et al. https://doi.org/10.1016/j.jaerosci.2023.106182.