

Feedback on the review by referee#2:

We thank Referee #2 for his/her comments, which we have addressed (The comments will be repeated in blue while the answers will follow in black.)

Q: "This paper seems overly long, and makes for tedious reading. An intercomparison between six units of the same instrument hardly seems publication-worthy. Far more interesting and useful is a comparison between different types of instruments, for example, the Boston College studies."

A: We agree that comparing the measurements of a new instrument against other independent measurements of the same quantity is a crucial step in assessing the accuracy of the new method. However, a thorough test of the reproducibility of a method by comparing multiple "identical" instruments is also an important characterization step (a method that is not reproducible might e.g. by chance provide "accurate" results in a comparison with independent measurements). Careful instrument characterisation may indeed be tedious work at times, but it is necessary for valuable scientific measurements. Our study focuses on the reproducibility of the SP2 measurements – as clearly stated in the manuscript title – but it also provides some comparison with independent methods. This kind of study fits the scope of the AMT journal well and the reader knows what to expect.

Q: "It should be noted that the second BC study (Cross et al., AS&T 2010) actually compared 3 SP2s against a CPMA (particle mass measurement), and stated that differences in response were merely due to a choice of detector gains."

A: The second Boston College study indeed probed BC with a wide range of instruments, including 3 SP2s. However, the paper by Cross et al. (2010) only reports - likely due to space restrictions in a more general overview paper such as this - very little results on SP2 comparisons:

- 1) A CPMA has been applied to calibrate the 3 SP2s with thermally denuded flame generated soot (FGS). Result: the three SP2s have different detector gains.
- 2) It has been shown for quasi-monodisperse FGS samples that changes of particle morphology have no influence on the SP2 BC mass measurement (using the same CPMA as a mass reference).
- 3) It has been shown for quasi-monodisperse FGS samples that coating of the BC cores with non-refractory matter has no influence the SP2 BC mass measurement (using the same CPMA as a mass reference).

This leaves many open questions such as e.g:

- How would the SP2s compare for e.g. diesel exhaust or ambient BC, which could potentially be different from the FGS applied in the Boston College study?
- How would the SP2s compare for polydisperse aerosol samples, i.e. has the signal-to-noise ratio of BC mass measured in individual particles an influence on the resulting BC size distributions?
- How would the SP2s compare with other methods? Many different methods were applied in the Boston College Study, however, Cross et al. (2010) compare the SP2 only against the CPMA, which was also used to calibrate the SP2s.
- How would the SP2s compare for the optical sizing?

- How would the SP2s compare for the coating thickness measurement?

- ...

Our study does certainly not address all possible questions but we believe that it contributes to improve the understanding of the precision and accuracy of the SP2 method regarding different data products.

Q: "As for scattering diameter estimates, a similar consideration (appropriate calibration) applies."

A: We don't consider it justified to conclude from the fact that the SP2 measures BC mass reliably that it also measures optical sizes reliably as two completely different measurement principles are applied for these two measurements. To our knowledge, there is no previous study available in literature that shows a comparison of optical sizing and coating thickness measurements between several SP2s or between an SP2 and other methods.

Q: "There are certain interesting sections in this manuscript – for example, the comparison of Aquadag and fullerene soot calibration-based estimated BC mass. But that follows the authors' earlier manuscripts, and as such, could just be published as a note, rather than a full-blown manuscript."

A: The result of our study that the sensitivity ratio of the SP2 between Aquadag and fullerene soot is essentially equal for 9 different SP2s has not been shown elsewhere. This is a key element in the argumentation chain put up in the earlier manuscripts (Baumgardner et al., 2012) that Aquadag calibrations can be recalculated to fullerene soot equivalent calibration curves.

Q: "A good portion of the current submission essentially boils down to: "Follow the manufacturer's instructions for setting up and maintaining your SP2." That is not science."

A: It is our decided opinion that it would not be science to simply follow the manufacturer's instructions for setting up and maintaining the SP2. Two examples are:

1) The manual states that size-selected Aquadag particles should be used for calibration of the incandescence detector (literature recommends fullerene soot calibration or a scaled Aquadag calibration; Baumgardner et al, 2012) but they don't provide a parametrisation or literature reference for calculating the particle mass from its mobility diameter.

2) Furthermore, the calibration procedure for the scattering detector is not even mentioned in the manual.

3) The manufacturer recommends the following maintenance (SP2 Operator Manual): "The only routine maintenance that the SP2 requires is to periodically refresh or replace the desiccant in the drying cartridge on the purge line..." and "An occasional particle zero check is also recommended..." and "If the SP2 is operated in a highly particle-laden environment, it may be necessary to check the calibration of the laminar flow element (LFE) on the sample inlet...". This is – to our opinion – insufficient to assure high-quality SP2 measurements (see Sect. 2.1.3 of our manuscript). Additional useful instructions for setup and maintenance of the SP2 are provided by the manufacturer, however, a checklist (or equivalent) giving clear instructions on all preparation steps and

tests required to achieve optimal SP2 performance and therefore high quality SP2 measurements is not available.

4) In addition, in our experience, the manufacturer's manual is a work in progress that is continually improved based on results such as ours in the literature.

Q: Page 3524:

Line 12: the verb is “incandesce”, not “incandescence”.

A: Done

Q: Line 13, minor quibble: Thinking of how the SP2 measures mass, any method other than gravimetry will be indirect (and nobody is going weigh individual BC particles, and even if that happened, there are interferences due to non-BC materials). That qualifier/description struck me as unnecessary.

A: The authors believe that this qualifier is still useful. For example particle mass analysers measure the mass(-to-charge) of individual particles directly. That is why they are useful for calibrating the SP2 (the only trouble there is to get pure BC particles with the appropriate chemical structure). Mass spectrometers also measure mass (-to-charge) directly, though it is the mass(-to-charge) of possibly fragmented ions and it is of course not always trivial to quantitatively relate the mass spectra to the original mass of the corresponding components in the aerosol sample.

Q: Lines 20-21: “several research groups” – who are these groups? What is “several” –5? 25? What modifications have they made? Are any of them included in this study?

A: This has already been clarified in response to a comment made by referee #1 (see below):

The following sentence was added in Sect. 2.1.1: “Only unmodified instruments were used in this study to compare the performance of different SP2 instruments.”

Section 5 and Fig. 13 contain data from three additional instruments from NOAA in Boulder. The following clarification has been made in Sect. 5 of the revised manuscript: “Figure 13 shows the ratio of the SP2’s sensitivity to Aquadag and fullerene soot in dependence of particle mass for the broadband incandescence detector of all SP2s involved in this study and three additional SP2s from the Earth System Research Laboratory at NOAA in Boulder. The NOAA instruments have a modified narrowband incandescence detector (see also Sect. 2.1.1), however, the broadband incandescence detector, which is used here, uses the same type of detector as the other 6 commercial instruments.”

Q: Line 24: “two elements APD” – it’s actually four elements, of which either two elements are used (e.g. Gao et al. 2007), or the four elements are wired to output two signals.

A: This is correct. However previous literature including the original paper by Gao et al. 2007 referred to this detector as the TEAPD. We believe that it is not useful to introduce an additional term here. Instead we added a reference to Gao et al. 2007, for those who are interested in learning more details about the multi-element APD.

Q: Page 3525:

Lines 14-15: “Both amplification gains can be varied while the high-to-low ratio remains constant” – this is not easy, as the gains are not linear settings. Further, this practice seems unnecessary.

A: The detector boards including the amplification stages are designed by the manufacturer in such a way that the amplification ratio between the high and low gain outputs is fixed, while the amplification gains can be adjusted by the user with a single adjustable resistor (any resistor adjustment changes the gain of either channel by the same factor). A fixed ratio – if it is well chosen – does actually make sense as it assures a constant overlapping range, where both the low and high gain outputs provide valuable data.

Q: Lines 24-25: “The low gain output:” – do the authors suggest that the high gain channels from the older SP2s are not the same as the high-gain channels from the C* SP2s?

A: The detector gains can be adjusted by the users (see previous comments). The C* SP2 of LGGE was adjusted (by the manufacturer) to almost maximal signal amplification. Therefore the high gain output saturated at relatively low BC mass, while the amplification of the low gain output was comparable to the amplification of the high gain output of the other SP2s. However, this study showed that variation of the amplification gain does not affect the quality of the SP2 measurements other than having an influence on the lower and upper detection limits.

Q: Section 2.1.3: The laser power adjustment on page 3527 is the only adjustment that may not seem standard manufacturer’s recommendation. Even then, the laser power is always maximized – not just with the pump laser current, but also with the pump laser temperature. I am not certain that the color temperature is dependent on laser power, except perhaps at very low laser power settings, where one runs into detection efficiency issues anyway – not advisable! All of this section can be deleted.

A: As mentioned earlier, the list of recommended SP2 adjustments required to achieve optimum performance of the SP2 is, in our opinion, useful.

More specifically, simply increasing the laser power to the maximum that can be achieved is insufficient to ensure sufficient laser power. The laser power should therefore be adjusted following the procedure described in Schwarz et al. 2010, which relies on the laser power dependence of the color ratio, in order to ensure that the calibration of the incandescence detector is independent of laser power and to avoid measuring with a laser power that is not sufficient for maximum detection efficiency of the small BC particles. This is the very point of having this item in the list of recommended SP2 adjustments.

Q: Page 3528, Line 18: That should be “NIOSH 5040”, not “NIOSH 940.” Further, the NIOSH 5040 does not recommend different temperature set points and residence times, either in itself: <http://www.cdc.gov/niosh/docs/2003-154/pdfs/5040.pdf> or in the reference: <http://www.cdc.gov/niosh/docs/2003-154/pdfs/chapter-q.pdf>

A: The NIOSH protocol number was corrected.

Q: Page 3532: Lines 4-5: Couldn’t the authors test the CAST soot with an APM/SP2 combination?

A: Indeed, we tried to involve a research group with a particle mass analyzer into this measurement campaign, however, we did not succeed.

Q: Page 3533, Lines 15-16: LEO stands for Leading Edge Optimization. That sentence does not mention “optimization”. Maybe the authors meant to say “using the Leading Edge Optimization (LEO) method”?

A: The acronym, defined in Gao et al., 2007, stands for “Leading edge only”. We slightly rearranged the sentence to:

“...from the leading edge only (LEO) of the scattering signal...”

Q: Page 3534, Lines 6-7: “the LEO fit was performed up to $2\mu\text{s}$ before:” – the justification for this choice is not clear, and perhaps even wrong. For example, with thickly coated BC, the “incandescent lag” – the gap between the scattering signal peak (after which the coating evaporation is obvious) and the incandescent peak is about $4\mu\text{s}$. The rise to incandescent peak is very quick. Here’s Gao et al. (2007):

“The chosen end point for the LEO fitting (end of thick line) must occur before the particle boiling point is reached. The choice of about $2.5\mu\text{s}$ before the coating starts vaporizing yields an amplification factor of approximately 30, consistent with the first method.”

So I don’t think the authors’ method for LEO is optimal.

A: Choosing the ideal limit for the LEO-fit is indeed a delicate business. It is a trade-off between improving the signal-to-noise and risking to underestimate the coating thickness due to partial evaporation. Gao et al. (2007) showed (their Fig. 11 and associated discussion) that choosing a limit between $\sim 3\text{-}5\%$ of the maximal laser intensity produced reliable and unperturbed coating thickness measurements for their particular laser intensity. We have chosen $2\mu\text{s}$ before the *earliest* incandescence as a LEO-fit limit in our data analysis – is this an optimal choice, given the fact that the time-lag between the scattering peak and the incandescence peak is in the order of $3\text{-}4\mu\text{s}$ for thickly coated particles?

The time when incandescence of a particle occurs depends on several factors. Higher laser power, larger BC core mass and thinner coating result in earlier incandescence, and linked to that earlier coating evaporation (“earlier” is here meant in the sense of absolute position in the laser beam). This shows that the optimal choice for the LEO-fit is linked to the time when incandescence occurs. Therefore it is not a good idea to use a fixed “standard” limit for the LEO-fit, instead it should be carefully determined for each data set and even BC core size range under investigation.

The earliest observed incandescence corresponds to a very small number fraction of very large BC cores. Choosing the LEO-fit limit $2\mu\text{s}$ before the earliest incandescence time, as we did in this study, is therefore ok for the vast majority of smaller BC cores (incandescing at a later time), which give the dominant contribution in terms of number. In order to show this, we followed the approach previously introduced by Gao et al. (2007) and varied the limit for the LEO-fit.

The LEO-fit was performed up to two different limits on the same ambient sample: once up to $2\mu\text{s}$ before the earliest incandescence, as applied in the discussion version of the manuscript, and a second time up to 1% of the maximum laser intensity (translating to $2.4\text{-}4.2\mu\text{s}$ before the earliest incandescence, depending on the instrument; performing the

LEO-fit to even lower limits would get increasingly unreliable). The BC coating thickness obtained with the two different limits was compared for each instrument. The figure below shows that the data points corresponding to individual particles essentially scatter about the 1:1-line. The histograms of the coating thickness (not shown) do not change significantly when choosing the lower limit for the LEO-fit. As a conclusion it can be said that our choice for the LEO-fit limit is in the appropriate range.

The following sentence was added to the manuscript in order to highlight this point (Sect. 3.3):

“Careful tests with more restrictive LEO-fit limits have been done to ensure that the above choice of LEO-fit limits is appropriate for all SP2s.”

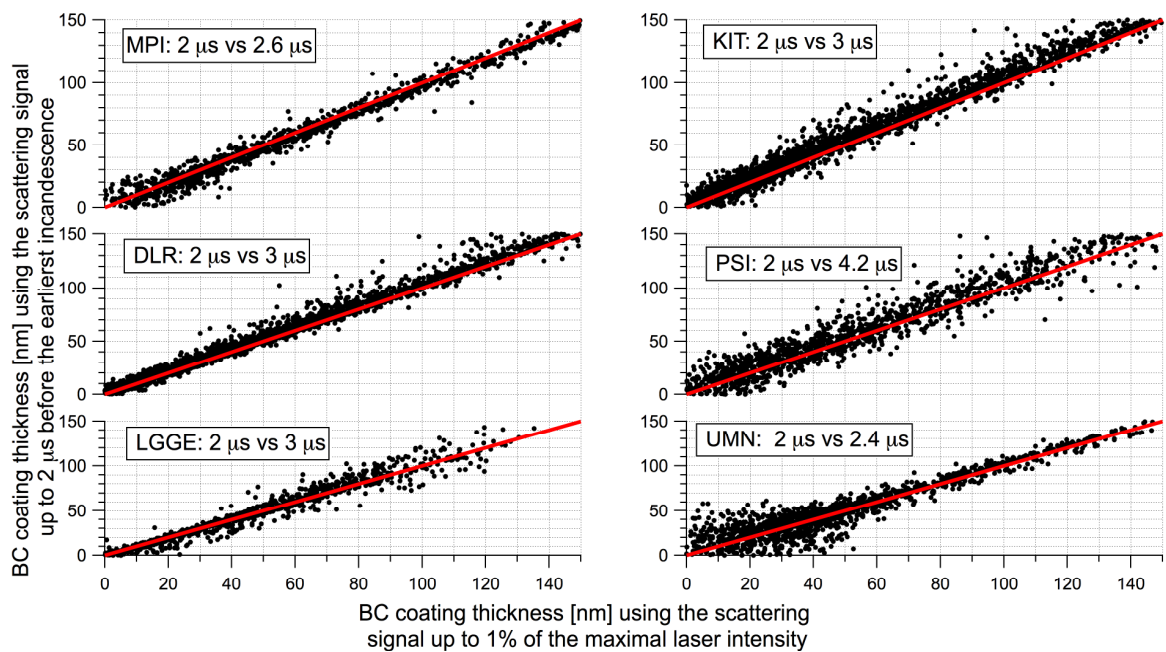


Figure 1: Sensitivity test of the BC coating thickness obtained from the LEO-fit to the amount of signal used in the fitting procedure.