Feedback to reviewer number 1:
We would like to thank Dr. McMeeking for his useful comments, which help clarify and improve the scientific quality of this paper.

Q: One of the main things that seems to be missing from the manuscript is a table summarizing the key statistical parameters describing the overall agreement of the instruments in the major categories (number concentration, mass concentration, size and coating/mixing state).
A: A table, summarizing the statistical parameters obtained during this study, was added to the manuscript (Table 5).

Q: In addition, more description of how the authors decided on the +/- XX % values reported in the manuscript is also needed. It isn't really clear where these values come from (they appear to be the range of minimum and maximum values) but in some cases there are clearly points outside this range (e.g., Fig. 7). Since the comparison of the SP2 outputs is a major focus of the paper, these numbers need to be defined more rigorously and summarized in a table. Right now they seem more like "ball park" estimates of the agreement.
A: The values reported throughout the manuscript and in the summarizing table are indeed obtained by taking the range between the minimum and the maximum values. This should be clear in the text and it is also stated in the caption of the new Table 5. Figure 7 indeed contains a few outliers outside the indicated repeatability range. However, the outliers at large BC core diameters are simply caused by limited counting statistics and would disappear if larger averaging times or size bins were chosen. The outliers below a BC mass equivalent diameter of 90nm are already discussed in the AMTD manuscript (beginning of Sect. 4.3): "The reproducibility of the BC mass determination was tested using diesel car exhaust particles. Figure 7 shows that the BC mass size distributions of diesel soot measured by all SP2s agree within ±10% above DMEV=90 nm. Below this diameter, the difference increases to ±20% at 70 nm. This increased uncertainty can again be explained by counting efficiency effects, but is probably also due to a higher calibration uncertainty at small BC mass."

Q: The abstract should state that the comparison values result from an analysis using a single analysis software package. Figure S1 shows good agreement between the results from two different software packages but I think it should be emphasized more strongly somewhere in the manuscript that the agreement seen here does not account to additional variability in output data that could result from differences in the analysis routines as well as potentially subjective choices for particle filtering made by different users.
A: This point was clarified in the abstract by adding the following sentence to the abstract:
"It has to be noted that the agreement observed here does not account for additional variability in output data that could result from the differences in the potentially subjective assumptions made by different SP2 users in the data processing."

And the following statement has been added to Section 3.1:

“The statistical parameters obtained here therefore reflect only the difference between instruments and it has to be noted that additional variability in output data could result from the differences in the potentially subjective assumptions made by different SP2 users in the data processing.”

Q: It might be helpful to quantify the lower detection limit of the SP2s tested in this study using a D50 type approach, where the lower detection limit is set to the mass at which 50% of the particles are detected. This is a common metric for specifying size ranges of impactors and would be a good way to quickly summarize the different lower limits as well as their variability from instrument to instrument.

A: Section 4.1.3 was changed in order to include the D50 values for each instrument:

“In order to accurately compare the different counting efficiencies, the threshold diameter (D50) at which the detection efficiency is 50% was determined for all SP2s by fitting a sigmoid curve to the data points (note: these diameter values represent mass equivalent diameters which are calculated from the BC mass assuming a bulk material density of 1800 kg m^{-3}).LGGE’s SP2 gets the lowest lower detection limit (LDL) with a D50 of 57 nm.”

and

”Considering only the well aligned instruments (Fig. 4; solid lines), MPI, UMN and KIT SP2s have similar counting efficiency (D_{50} ~67 nm) while PSI’s SP2 is a little less efficient (D_{50}~80 nm) followed by DLR’s instrument (D_{50} ~90 nm).”

Q: The comparisons to non-SP2 techniques should either be described in more detail or else omitted. Additional information on the size range of aerosol sampled by the optical and OC/EC instruments needs to be given and the discussion expanded. For example, in current form it is difficult to evaluate the large MAC observed for the CAST soot referred to in section 4.3. Some of this difference is likely due to the contributions by particles outside of the SP2 size range. Comparisons of the SP2 to other BC and EC measurement methods are needed, but given that the focus of the manuscript is on an inter-comparison of SP2 instruments it may be better to omit the comparisons in this work and hopefully address them in a more detailed manuscript in the future.

A: The authors agree with the fact that the main focus of the manuscript is on the inter-comparison of the SP2 instruments. However, the question of the absolute measurement accuracy is also included with comparing the SP2 number concentration and number size distribution measurements against independent CPC and SMPS measurements. Furthermore, the discussion about the effect of using different calibration materials touches the question of the accuracy of BC mass measurements. Therefore we also retain the comparison of SP2 measurements with the independent EC mass measurement by the Sunset thermo-optical OC/EC instrument. The following clarification has been added to the revised manuscript: “The comparison of the SP2 BC mass with the Sunset thermo-optical EC and the calculated MAC values should not be affected by the limited detection range of the SP2, as the mass size distribution of the CAST soot used for this comparison was well within the detection range of the SP2.”

Specific

Q: Page 3524, lines 17-21: Is this sentence referring to the SP2 research groups at large or just those involved in the study? If any of the instruments discussed here were modified this should be noted.

A: 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, is identical to the other 6 commercial instruments."

Q:Page 3526, 17-21: would be helpful to have specific sizes for the "larger particles" referred to here.

A: Specific size will depend on the instrument configuration and cannot therefore be given here as a general rule. We added the following general sentence:

"larger particles (in a size range depending on the instrument configuration).

Q: Page 3527, 24-26: was there any reason for the lower flow rate of the MPI instrument? Was this to test sensitivity of the instrument response to varying flow rate?

A: No, this was not intentional. The calibration of the internal sample flow measurement was wrong, resulting in a lower flow than the nominal value (which was set to 0.12 L/min for all SP2s). The sample flow rate check was not performed at the very beginning of the measurement campaign and therefore the MPI SP2 happened to measure at a different sample flow rate.

Q: Page 3528, 27: should add that the interference occurs on filters.

A: This sentence reads now: "Basically, it uses a power-modulated frequency-doubled Nd:YAG laser (lambda=532 nm), in an optical resonator equipped with a microphone, for a direct airborne measurement of the aerosol absorption coefficient. This method is unaffected by light scattering from the filter matrix and the collected aerosol as it would occur with a filter-based measurement."

Q: page 3532, 24: might as well state the solid angle here as well to be complete

Done (Beginning of Sect. 3.3).

Q: page 3533, 25-28: please state the refractive index values used in this calculation (same as Schwarz et al. 2008?)

A: Several sections were modified in order to include the two refractive indices used during this study (for ambient BC and CAST soot):

The discussion was moved from Sect. 4.5 into Sect. 3.3 where the calculation of the BC core RI is discussed more generally. The sentence was also modified in order to include

the ambient BC core RI used and reads now: “Consistency between BC mass measurement and optical sizing of the BC core was achieved for the CAST soot with using an RI of n=1.9+0.8i. This value is within the likely RI range of light absorbing carbon reported by Bond and Bergstrom (2006). Using an RI of n=2.26+1.26i resulted in consistent results for the ambient BC particles, which is in agreement with literature (Moteki et al., 2010). In both cases, an RI of n=1.5 was used for the coating matter. This RI has previously been used for similar coating thickness analyses of ambient particles (here fits probably the Schwarz 2008 reference) and it is also representative of aliphatic hydrocarbons and polycyclic aromatic hydrocarbons (PAH; Sutherland et al., 1994), which are major components of the OC in fresh CAST soot (Slowik et al. 2007). The composition of the organic coatings of CAST soot changes during the addition of SOA coatings. The quantitative comparison between the coating thickness measurement and the independent OC volume fraction measurement by the Sunset instrument (Sect. 4.5 and Table 3) is done for fresh CAST soot. Therefore we use the RI representing the OC coating of fresh CAST soot for the coating thickness analysis of this experiment (Sect. 4.5 and Fig. 12).”
The refractive index assumptions are also added to the figure captions.

Q: page 3536 (19), 22: "almost stable" please be more specific (i.e., percentage of drift from initial value or something similar...)
A: The sentence reads now: “The integrated laser power was almost stable for the SP2s from MPI, UMN, LGGE and PSI with respectively a drift of 6.3%, 0.8%, 1.6% and 4.3%. KIT’s SP2 laser power increased dramatically after the mode aperture realignment (done on 29 November 2010). DLR’s SP2 laser power gradually dropped (with a drift of 27%)”

Q: page 3538, 15: suggest changing "insure" to "ensure"
A: Done

Q: page 3538, 17-18: "...an uncertainty of 20% in scattering amplitude still provides decently accurate optical sizing." It would be helpful to have some firm numbers here for the typical limits of the scattering size range, such as +/- XX nm for a 150 nm particle at the lower size range and +/- XX nm for a 400 nm particle at the upper limit.
A: The authors believe that this amount of details is beyond the scope of this paper. The overall optical sizing uncertainties depend on many factors such as the detector calibration uncertainty, raw data processing, and assumptions on the optical properties made in the data analysis. Currently we have a manuscript in preparation that discusses a comprehensive concept for the calibration and analysis of the SP2 scattering signals including consistency checks against the incandescence measurement, mobility sizing and detailed sensitivity analyses (Gysel, 2012, in preparation).

Q: page 3539, 4-7: It would be helpful to expand this section slightly. How much did the filtering thresholds vary from instrument and how sensitive are some of the comparisons to the choice of filtering value?
A: This statement is not about any filtering in the processing of recorded raw data. Instead it is about the fact that the signals of small particles are only stored if the triggering threshold in the data acquisition software is set close enough to the baseline.

The choice of the triggering threshold only affects the lower detection limit, as already stated in the manuscript. Any other comparison of quantities which do not have a substantial contribution from signals right at the lower detection limit is not affected by the choice of the triggering threshold. In this study we generally tried to keep the triggering thresholds as close as possible to the baseline in order to keep the lower detection limit as low as possible. We prefer not to add further discussion about the triggering threshold to the manuscript, as this would dilute the main messages.

Q: page 3541, line 11-14: please clarify if this refers to CAST soot before or after coating with a-pinene SOA.
A: Done

Q: Text in some figures (e.g., fig 14) is too small.
A: Done