Answers to referee 1

Comparison of co–located rBC and EC mass concentration measurements during field campaigns at several European sites

We thank the reviewer for the timely and constructive review of our manuscript, particularly during these difficult corona times. Please find below reviewer comments repeated in black text and our responses in blue text.

1) While the authors give a relatively good overview of the specific problems associated with EC/BC measurements, in highlighting the general background (i.e. the nature of light-absorbing carbon in the atmosphere) they fail to account for established prior knowledge in the field. They seem to cherry-pick information from the literature, ignoring important findings that would be needed for better understanding the complexity of the problem and the difficulties with the selected methodologies.

   The level of methodological detail to be provided in such a manuscript comparing two existing methods remains a subjective choice as the useful level always depends on the reader's prior knowledge on one and/or the other method. Based on the answers given below in response to the referee's specific comments, it appears that the impression of "cherry-picking" was partially caused by keeping artefacts of lesser importance in Sect. 2, while addressing the more relevant ones in Sect. 3. We aim at clarifying these aspects with the revised manuscript.

   For example, the issue of light absorbing carbon continuum is referred to in the manuscript as being newly discovered (supported by two recent papers of the authors themselves, line 80-85), though it has been on the agenda of aerosol science for nearly 20 years.

   It was by no means our intention to imply "newly discovered" with our statement "refined classification", nor was it sold like that in Corbin et al. (2019). The submitted text referenced by the reviewer is:

   > *Recently, Corbin et al. (2019) proposed a refined classification of light–absorbing carbonaceous PM into four classes: soot–BC, char BC, tar brown carbon and soluble brown carbon, and they provided an overview of the respective physico–chemical properties. This refined classification provides a useful framework in describing the responses of TOA and LII. For example tar brown carbon, an amorphous form of carbon, is sufficiently refractory to contribute to EC mass, whereas it is not sufficiently refractory to cause substantial interference in rBC (Corbin and Gysel–Beer, 2019).*

   Here, and in our earlier work, we emphasize the concept of a refined classification with respect to measurements. Earlier excellent work is extensively cited in our recent papers. We agree with the reviewer that earlier papers such as that by Bond (2001) have already raised this issue. However, the terms tar and brown carbon were not used by such earlier work, yet have become very popular in recent years (e.g. the next comment by this reviewer). Therefore, in our discussion of brown carbon, we have cited our recent work rather than the earlier work.

   We agree that we missed the opportunity to cite original work such as Bond (2001) at the appropriate place, and we will add a citation to that landmark paper in the revised manuscript.

Bond, T.C. (2001). Spectral dependence of visible light absorption by carbonaceous particles emitted from coal combustion. *Geophys. Res. Lett.*, 28(21):4075–4078.

In general, more focus should be put on this issue, including brown carbon, as this may greatly affect differences between the results of the measurements. Some statements such as 'brown carbon absorbs much less than EC at the red wavelength (λ = 635 nm)' (line 211-213) are largely outdated and ignore more recent findings that there are actually two types of brown carbon in aerosol, the strongly absorbing tar balls (carbon spheres) and the weakly absorbing ones (organic chromophores), the former have also been shown to absorb in the near infrared (see e.g Alexander et al. 2008, Saleh et al., 2014, Hoffer et al., 2017).

We agree that tar brown carbon does absorb light also at red and NIR wavelengths. Our statement was not clear enough, and was intended to be 'brown carbon absorbs much less **per unit mass** than EC at the red wavelength of the laser…'. We will add the text in bold.

2) My another major concern is that even if the authors primarily make 'lump statistics' on all their measurements data, their individual measurement setup is different at each site (e.g. different cutoffs, dryers, SP2 instruments, etc.). In measuring such small-sized and adhesive species even the type and length of the tubing may introduce significant uncertainties due to wall losses, these are likely also different in the measurement sites but not reported here.

Diffusion losses are often an important source of error for particle number related quantities, whereas impact on particle mass related quantities is typically smaller. Indeed, diffusion losses were estimated to be less than 10% across the size range of the SP2 (see Sect. 3.3.1), mainly associated with using a dryer in the sampling lines (see Sect. S4).

3) My fundamental question is the following: if we take into account these uncertainties (or biases?) and the other existing uncertainties (and potential biases) correctly evaluated and reported in the manuscript (I counted 8 significant sources of uncertainties but these may not be all), correctly apply the rules of error propagation, will the overall 8 % difference between the two methods be statistically significant at all? My informed guess is that it will not.

So contrary to the key statement in the manuscript that 'median ratio between observed rBC and EC mass concentrations was 0.92', a more realistic statement would be that 'the two independent methods are indistinguishable within the limits of inherent uncertainties'. Overall, the general impressions from the discussions and conclusions tacitly support this latter statement as the inexplicable variations (to either directions) of campaign-wise data do not reveal any systematic difference between the two methods. This is particularly true since the individual campaigns at the four sites were markedly different in duration (14, 21, 24, 30, 54 days according to my calculations), so the 'mean data' reported in this manuscript refer only to this particular combination. Should the durations of the individual campaigns be different, the overall finding would have been very much different, at least within the limits of extremes. Campaign-wise discussions would have made more sense than calculating a 'European average' value with this fixed setup (which definitely does not exist).

The quote made at the beginning only includes a half sentence taken from the abstract. We always provide the geometric standard deviation along with the geometric mean ratio, also in the abstract, both for the overall and the campaign-wise data sets. For this 0.92 ratio, the GSD covering 68% of

the data points was a factor of 1.5. Hence, we believe that the abstract taken as a whole conveys the same message the reviewer is suggesting, and in a quantitative manner.
We also provide recommendations on how to design future experiments towards achieving tighter error margins (e.g. at the end of Sect. 3.3.1 and in the conclusions): "For future intercomparison studies, it is important to constrain the upper cut–off and potential inlet losses of both methods in such a manner that these can be excluded as a source of discrepancy."

4) My last major concern is that the authors do not exploit the possibilities of elaborating differences within the individual campaigns (though the campaigns are not too lengthy in themselves). For example, dependence of measured results on trajectory directions at the Melpitz site should have been elaborated to verify the hypothesized effect of coal burning from Eastern Europe. Although these evaluations would not involve statistical analyses, they might still be useful to imply some of the effects that are hypothesized in the manuscript.

The referee can be assured that we spent plenty of effort trying to understand any potential reason for systematic deviations between the two methods, be it between different campaigns or within a campaign for different air mass origin or variable contributions of sources. This includes:

- Differences in upper size cut-off between the two methods.
- Undetected rBC mass below the lower size cut-off.
- Source specific sensitivity or artefacts of one or both methods.
- Known artefacts or interferences from e.g. tar brown carbon or instrument operation outside parameter range ensuring optimum performance.

Interpretation of these exercises turned out to be a challenging trade-off between providing as much insight as possible while avoiding over-interpretation. As an example: The back trajectory analysis was done for the Melpitz winter example raised by the referee, and we discussed in the last paragraph of Sect. 3.1.1 how coal combustion influence from Eastern Europe likely contributed to a shift to larger BC core sizes and hence reduced missing rBC mass below the lower LOQ of the SP2. At the same time, this shift towards larger BC particle sizes resulted in above-average fraction of total EC in PM2.5 found between 1.0 and 2.5 µm aerodynamic diameter (see Sect. 3.3.1), thereby increasing comparison uncertainties associated with different upper cut-off.