

The paper describes a new type of analyzer that can be coupled with the AE33 Aethalometer to provide Total Carbon measurements and estimate equivalent Organic and Elemental carbon concentrations in particulate matter deposited on a filter. While a field analyzer offering OC/EC/TC measurements has already been available in the market, this new instrument incorporates several novelties, mostly in design, like the use of a dual sampling channel, application of zero air as carrier gas and durability of materials. The instrument and method can potentially benefit AE33 users and help them obtain additional TC, eOC and eEC results. It can further compliment the offline OC/EC reference method with high-time resolution measurements. As with all new analyzers, it should go through type testing and validation, which is the main weakness of the current paper since this is partially dealt with.

Comments:

As mentioned on the Title, page 1 line 16, page 4 line 1 and elsewhere the authors state that this is a new or newly developed method. Nevertheless, a method bearing great similarities has been already described in the past (Bauer et al, 2009 and references therein). The paper describes an analyser of a different manufacturer that “also determines optical OC and optical EC by monitoring the laser transmission data through the quartz filter”, “Total carbon (TC) is determined using the thermal-optical method, and then optical OC is deduced by subtracting optical EC from TC (optical OC = TC - optical EC)”. Where optical OC and EC would be simply a different terminology given for eOC and eEC used in the current paper. How would the authors comment on the method similarities of the two studies and the suitability of the description “new” for the method?

While the terms of OC, EC and eOC, eEC are clearly defined, their use in the text overlaps and is occasionally confused. Proper terminology should be consistently used in order to avoid any misinterpretations by readers. For example, the abstract mentions in lines 22-23 that this new application can result in high time resolution determination of organic and elemental carbon while in reality it provides an estimation of eOC and eEC values. Another example would be in section 3.6: eOC and eEC should be used instead of OC and EC. Also applicable in all graphs.

NDIR detectors, similarly to the one in the current application, may deteriorate in performance in long term and show a drift in their baseline. Since there is no application of an internal standard calibration or a span check, have the authors evaluated how often would an external standard calibration be required? Would there be any NDIR detector related maintenance needs, e.g. source replacement, and in what frequency would that be required?

The last paragraph of section 2.3 describes tests performed on the denuder efficiency but it seems that results are not included in the paper. Page 7, line 21 also refers to TC data “(see below”, which are not presented later on in the text. Related to the denuder efficiency, there is clear evidence of a positive artefact for eOC concentrations below $8\mu\text{g}/\text{m}^3$, visible in Figure 8 (OC vs eOC) as well as in Figure 11 (b) (OM vs eOC). Further there are signs of a negative artefact for higher concentrations, based on the same graphs, suggesting reduced combustion efficiency. The later would be more profound if there would be an addition of second denuder monolith as suggested in page 12 lines 25-27 or if a correction for the positive artefact would be applied. Should the user then consider the use of 2 correction factors (b)

related to the concentration levels measured? Or would there be any other suggestion to overcome these issues?

Section 3 refers to EN 16450:2017 regarding the orthogonal regression analysis on the 31 daily measurements between the new instrument (candidate method) and 2 independent laboratories (reference method). A proper application of EN16450:2017 would require a minimum of 40 valid data pairs with the further requirement of 2 candidate applications for each type testing application. EN16450:2017 further describes requirements related to the number of locations and the concentration range of data points. The use of just one candidate method limits the conclusions on performance consistency between identical instruments and restricts the candidate method uncertainty calculations. It seems that quite a significant part of section 3 discusses the comparability between the two reference method applications, which is a topic thoroughly documented elsewhere in the literature (Inter-comparison exercises publications are included in the reference list of the current paper). A more relevant approach would have included and compared two new instruments in parallel measurements. How would the authors comment on the approach applied and the data suitability?

The uncertainty limit value of $2.00 \mu\text{g}/\text{m}^3$ mentioned in page 10 line 13 originates from calculations of PM reference methods where limit values of $30 \mu\text{g}/\text{m}^2$ or more apply. Would that be directly applicable for TC method and concentrations? Further section 3.3 describes method uncertainty calculations based on the NDIR detector response. Have the authors considered additional sources of uncertainty to be included in the uncertainty budget, e.g. use of denuder, zero air carrier gas, ambient temperature and pressure variations?

High volume samples were collected for analysis by the two laboratories. Is there information available on the type of the filters used? Was the homogeneity of the samples estimated for this study? Did the laboratories analyze the samples in triplicate or duplicate and were there standard external solutions analyzed like in most of the comparison exercises referenced in the text? Standard solutions may provide an insight if the difference between the TC results of the 2 laboratories was a result of calibration deviations.

A common practice in comparison exercises is following identical procedures on filter handling, transport and storage for all participants. In the current case filters were first analyzed by the ARSO laboratory and then shipped to IGE for further analysis. Even though the authors mention that “sampling, transport and storage of the filters was done according to the EN 16909:2017” IGE received the filters for analysis at a later period, after additional transport, handling and storage. Could that have contributed to the small uncertainty observed between the reference methods? Wouldn't an approach of dividing the samples and shipping in both labs in parallel have resulted in improved comparability of the two laboratories?

Table 3 provides with a wide range of b factors and the recommendation right above is: “the determination needs to be performed for each location and with filters sampled over the time-period of interest”. Considering that the b factor is location and season specific, could the authors suggest a typical coverage period range with OC/EC offline analysis in parallel? How often would the b factor have to be re-evaluated per location?

The calculated b factor for this study (0.44) is the lowest among EUSAAR_2 users of the literature listed in Table 3. Further the slope from the OM – eOC comparison of (Figure 11,(b)) is 1.82, and would have

been even higher when considering the high negative intercept. Following the ranges provided by the literature in page 17, lines 14-15, these slopes fit better a rural site rather than an urban environment. This comes in contradiction with the characteristics of the selected site which is influenced from traffic emissions. If the estimated b factor was within the literature range that would have further resulted in a lower OM - eOC slope and would fit the literature range better. How would the authors comment on these observed differences compared to the literature?

Following Graph 7, it seems that in the first half of the campaign BC is overestimated while on the second half underestimated, compared to EC. Would there be any interpretations for this observation? Further, around the 5th of March eOC and TC from the TCA08 configuration are overestimated and bBC is underestimated, significantly more from the rest of the data points. Would there be any justification by the authors?

Page 15, lines 19-21, suggest that the intercept due to the eOC positive artefact can be neglected based on the comparability with the OC intercept of the two independent laboratory measurements. Nevertheless, as discussed earlier in the text, the eOC intercept is systematic and attributed to the denuder performance. It should also be noted that the eOC concentrations are compared to the average values between the two laboratories. It would be more appropriate if the artefacts observed for a new instrument would not be overlooked but rather investigated further and be dealt with. The use of two candidate method analyzers and an extended data set would be required for in-depth analysis. The above comes also in contradiction with the conclusion of page 20, line 12: "the correlation analysis showed very high agreement between eOC and eEC to the EC and OC". It seems that there is room for improvement in agreement once the artefact issues are resolved.

Technical corrections:

Page 2, line 26: tTe.

Page 15, line 4: It is not clear on which comparison the authors refer.

Page 17, line 21: Fig 10(b) to Fig 11(b).

Page 18: Figure 11 (a) misses the dashed trendline.

References

Jace J. Bauer, Xiao-Ying Yu, Robert Cary, Nels Laulainen & Carl Berkowitz (2009) Characterization of the Sunset Semi-Continuous Carbon Aerosol Analyzer, *Journal of the Air & Waste Management Association*, 59:7, 826-833, DOI: 10.3155/1047-3289.59.7.826