Point-to-point response to Referee #2 (RC2)

We are grateful for yours comments on how to strengthen our manuscript. Please find attached revised manuscript and see our point-to-point response to your comments below:

1. **Introduction: The intro part mainly discussed the definitions of different carbon fractions and various protocols for measuring the carbon fractions. However, some of the content is repetitive (e.g. it was discussed in Page 2 Line 26 to Page 3 Line 2 that a few factors will influence the OC-EC split while similar points were mentioned again in Page 3 Lines 9–27). The authors also listed out three protocols that have been widely used in different regions of the world (Page 2 Lines 16–24). Since in the following sections the online data were compared with the offline data obtained from the EUSAAR2 protocol, readers might expect to see more discussions on the specific differences among the protocols (EUSAAR2 vs. IMPROVE_A, EUSAAR2 vs. NIOSH).**

   We believe that discussion on the specific difference among protocols is beyond the scope of this paper and was very well described in Karanasiou et al. (2015). We focus on operational definition aspects and harmonization between different experimental approaches.

   Following sentence was added to the manuscript (page 2, lines 24-25):

   - This protocol has recently became part of the European standard for the determination of OC-EC in PM2.5 samples (EN 16909:2017, 2017). Detailed discussion on the specific difference among protocols can be found elsewhere (Cavalli et al., 2010; Karanasiou et al., 2015).

2. **Section 2.2: How is the performance of external calibration (using sucrose or KHP or other chemicals) by TCA08? What is the maximum carbon concentration tested?**

   External calibration of TCA08 was performed with punches of ambient filters with known TC content. This is done to simplify the calibration procedure on the field and to achieve better calibration accuracy (ambient filter includes mixture of different EC and OC fractions). Carbon calibration constant is defined with the slope between measured integrated pulse of CO_2 by TCA08 and known TC mass content of the filter punch (Figure 1). The maximum carbon content tested in the TCA08 was up to 100 µg of TC.

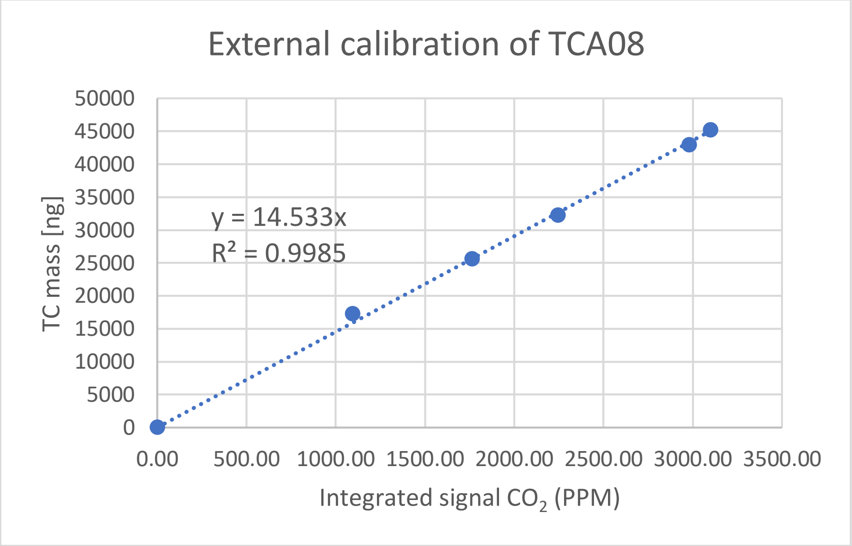

Figure 1 External calibration of TCA08 - determination of carbon calibration constant

3. **Section 3.3: What is the "carbon calibration factor"? How is it derived? What's the loading effect compensation algorithm used for treating AE33 data in this study? Will different algorithms introduce uncertainties?**

Carbon calibration factor mentioned in paragraph in section 3.3 is carbon calibration constant C_{carb} from Eq. 4. It is defined as the slope between TC mass on the punch of ambient filter with known TC content and the integrated value of the CO2 signal measured by the NDIR detector (Figure 1).

Carbon calibration factor was changed to carbon calibration constant (page 14, line 5):

- The uncertainty u_{TCA} associated with the TC data from the TCA08 includes individual uncertainty sources of the carbon calibration constant C_{carb}; the uncertainty of the analytic flow measurement; and the uncertainty of the signal and blank CO_2 peak measurement (Eq. 4).

Loading effect compensation algorithm used for treating AE33 data in this study is an real time "dual spot" loading compensation (Drinovec et al., 2015, 2017). The inlet air stream of the AE33 is split, and the sample is collected on two filter spots concurrently. The flow through each of the two spots is different, so the loading rates on the respective sample spots are different. Different loading rates cause the accumulation of the sample to be different between the two spots, resulting in a different magnitude of filter loading effect (FLE) between the spots. Measurement of FLE enables the compensation of the data – using the parametrization described in Drinovec et al. (2015), the compensation parameter *k* can be derived. We believe this method is the most appropriate one as it measures loading effect with the same time resolution as measuring black carbon concentrations and it does not make any assumptions. Using different compensation algorithms will introduce uncertainties based on the assumptions they use.

A following sentence was added to the manuscript (page 14, line 26):

- Figure 7 shows the regression of the off-line thermo-optical analysis of samples for EC (from the ARSO and IGE laboratories, using the EUSAAR_2 protocol) with the 24-hour averaged BC (Aethalometer data) obtained during the field campaign period. An AE33 integrated “dual spot” real-time loading compensation algorithm was used for BC data treatment (Drinovec et al., 2015).

4. **Section 3.4: It can be seen from this work and from previous literature data (Table 3) that the relationship between BC and EC is location dependent. If the aerosol composition at a certain location has a very clear temporal variation pattern, the parameter b could be sensitive to the sampling time period as well. Since the PM monitoring networks usually adopt the filter-based sampling approach followed by offline laboratory analysis and the historical dataset was very likely obtained from offline measurement, do the authors suggest that every time the online TC-BC system is deployed to one sampling location, the online-offline comparison needs to be conducted to derive the b value so that the measured data can be compared to other dataset?**

We agree with the reviewer. The sampling time period for offline filters can affect the b factor as well. We followed a common practice for sampling period; sampling start time was at 00.00 am and sampling stop time was at23.55 pm each day. During 5-minute idle period, the sampler automatically stored sampled filter and replaced it with the new one.

The proportionality parameter *b* is an effective value with a local and a regional component. Usually, the local contribution concentrations are dominant and the local BC and EC contributions dominate the relationship. The differences in *b* values presented in Table 3 show, that there is a big variation between different rural/regional background sites, and also between the urban sites. This is the reason why similar offline-to-online intercomparison is recommended for every new background site or site with strong mixture of local and regional contribution. The time period of intercomparison should cover seasonal variations in *b* values, for example 2-3 weeks each season. The re-evaluation intercomparison campaign for the certain location should be done if significant changes in the BC emission inventory is expected (traffic or wood burning restrictions, *etc.*). For sites with dominant traffic contribution, where the *b* factor mostly depends on the properties of the vehicle in the fleet, the intercomparison measurements will result in similar *b* values unless a significant fleet change occurs.

Following sentences were added to the manuscript:

- (page 8, lines 35-37)
  The TCA08 was operated on a 1-hour time-base, sampling $PM_{2.5}$ fraction at 16.7 LPM; co-located with a Model AE33 Aethalometer measuring Black Carbon aerosols in $PM_{2.5}$ on a 1-minute time-base at 5 LPM. At the same location, 24-hour $PM_{2.5}$ filter samples were collected in parallel with a Digitel high volume sampler for OC-EC offline analysis at two different laboratories; the Slovenian Environmental Agency (ARSO, Ljubljana, Slovenia), and IGE (Grenoble, France) both using the Sunset offline OC-EC analyzer with the EUSAAR_2 thermal protocol. Sampling start time was at 00.00 am and sampling stop time was at 23.55 pm each day. During 5-minute idle period, the sampler automatically stored sampled filter and replaced it with a new one. Additionally, non-refractory organic matter (OM) measurements were also performed during the campaign with an ACSM (Aerodyne, Billerica, MA; Ng et al., 2011) on a 29-30 min time-base to derive high-time resolution

measurements of the OM-to-OC ratio. The ACSM, equipped with a PM$_1$ aerodynamic lens, was sampling through a PM$_1$ sharp cut cyclone (SCC 1.197, BGI Inc.) at a flow rate of 3 LPM yielding a particle cut off diameter of roughly 3 µm. Furthermore, the sample was driven through a Nafion dryer, upstream the instrument inlet, keeping the sample relative humidity below 40% throughout the campaign. The chemical composition dependent collection efficiency of the instrument was determined according to Middlebrook et al., 2012. Due to variability in the ACSM time-base, we gathered the data into 3h averages. All of the instruments were checked regularly and operated without interruption throughout the campaign. No data were selectively removed from the results presented in the following.

- (page 15., lines 11-18)
  The proportionality parameter *b* (Eq. 3) is compared with values taken from the literature in Table 3. These values depend on the location, the nature of the aerosol, and the thermal protocol used for analysis. The value of 0.44 which we determined in this study for an urban background site is slightly lower than values for other urban and urban background sites using EUSAAR 2 thermal protocol, and considerably lower than the values for rural sites. The proportionality parameter *b* is an effective value that features a local and a regional contribution of BC and EC. Usually, the local contribution to concentrations is dominant and the local BC and EC contributions dominate the relationship. The differences in *b* values presented in Table 3 show, that there is a big variation between different rural/regional background sites, and also between the urban sites. This is the reason why similar offline-to-online intercomparison is recommended for every new background site or site with strong mixture of local and regional contribution. The time period of the intercomparison should cover seasonal variations in *b* values, for example 2-3 weeks each season. The re-evaluation intercomparison campaign for the certain location should be done if significant changes in the BC emission inventory is expected (traffic or wood burning restrictions, *etc.*) For sites with dominant traffic contribution, where the b factor mostly depends on the properties of the vehicle in the fleet, the intercomparison measurements will result in similar *b* values unless a significant fleet change occurs.