

We thank the referees for their comments, which helped improving the quality of our manuscript. A point by point response (in black) to the reviewers' comments (in blue) will follow. Changes in the text are indicated in in *black italics*.

Reviewer 1:

Organic aerosol (OA) is a major component of the total aerosol. Thus, understanding the sources of OA is crucial for controlling aerosol pollution. In the past decade, the development of the aerosol mass spectrometer allows the direct quantification of organic aerosol mass, and source apportionment can be done by further applying PMF on such data set. The main weakness of AMS is that the hard ionization technique causes significant fragmentation, so that some sources are difficult to separate, such as secondary formation from anthropogenic or biogenic origins. To overcome this drawback, the authors, for the first time, explored the usage of LDI technique on offline samples, which causes less fragmentation and may give new insights of OA sources. The authors developed novel procedures for instrument calibration and uncertainty assessment, generated reference spectra for PMF from both specific environment (in the tunnel for traffic sources) and the lab (for wood burning sources), and carefully evaluated the results by comparing to the results from AMS technique. By doing so, the authors reported both the advantages and some potential issues of this technique, which makes this work easy to understand and follow. Although the ionization efficiency and the "matrix effect" of LDI technique remains poorly understood, which weakened the interpretability of some PMF factors, this work provided several new insights in OA sources and should be published in ACP after the following comments are addressed.

Minor comments:

Page 4, Line 16-17. "The observed ions are typically still fragments: : :". Some of the spectra (e.g. Fig.3d) reported in this work contains some high-mass peaks, which might be complete molecules. This work will be strengthened if you can identify them.

We agree with the reviewer that the identification of ions in the recorded mass spectra would greatly help gaining a deeper understanding of sources and processes but also of the measurement technique itself. In the complex mixtures that we analyze, different ions have plausibly a significant contribution to the signal at a single nominal mass, especially given the size of the ions that we observe. With the mass resolution and mass accuracy of the m/z calibration an identification of the contributions of different ions is not possible unambiguously. Stark et al. (2015) developed an approach to estimate the carbon oxidation state using the mass-to-charge ratio of a peak and its mass defect (difference to closest integer) for mass spectra with limited mass resolution. With our instrumental approach a similar approach cannot be applied because of the poor mass accuracy of the m/z calibration.

We have adapted the text according to the reviewer's suggestion:

"... With this accuracy and resolution, neither distinguishing different ions at the same nominal mass nor estimating properties such as O/C and H/C for a nominal mass following Stark et al. (2016) was possible for the quartz-filter-LDI-MS measurements. Instead, the spectra were integrated to UMR sticks. ..."

As reviewer 2 mentions, the prominent presence of odd masses might hint to a rather prominent fraction of signal attributable to fragments as reviewer 2 mentioned. However, it is currently impossible to distinguish odd mass molecular contributors from fragments. According to Samburova et al. (2005a), fragmentation in this type of instrument is negligible. The mass spectral signatures acquired in this study are similar to the ones in Samburova et al. (2005a) and the lower laser energy applied in this study is lower than in the former study. This suggests that also during our measurements fragmentation was not a prominent process.

We added this discussion to the manuscript:

“ ...

The presence of considerable signal at odd masses might indicate that significant fragmentation occurs during desorption and ionization of the organic matter in the LDI-MS (Fig. 1a, 3, 4). However, Samburova et al. (2005a) suggested that fragmentation in this instrument is negligible. The mass spectral signatures acquired in this study are similar to the ones in Samburova et al. (2005a) and the lower laser energy applied in this study is lower than in the former study. This suggests that also during our measurements fragmentation was not a prominent process. Overall, the extent of fragmentation remains unclear.

...”

Page 5, Line 5-20 and Page 7, Line 5-10. The usage of AgNO₃ seems unnecessary to me. First, using UMR data does not require a high-quality mass calibration. Second, you mentioned that AgNO₃ may cause change in ionization physics, which might even hurt the mass calibration. If there are other important reasons for using AgNO₃, please specify. Otherwise, I think it is important to mention that spiking AgNO₃ should be skipped in future works. Following this comment, the mass calibration accuracy (Fig.3c) is very unstable and too low under the mass resolution of 1000-2000. Do you have explanations?

The m/z calibration parameters in our LDI-MS measurements were highly variable between different samples. Therefore, to determine unit mass resolution integration regions it was necessary to perform an m/z calibration on every single sample.

In other mass spectrometers, some peaks are always observed, independent of the sample, and can therefore be used for an m/z calibration and/or to assess the performance of an initial m/z calibration: E.g. in the Aerodyne AMS the ions N₂⁺, CO₂⁺, W⁺, etc. are present under most conditions. As another example, in chemical ionization mass spectrometry m/z calibration is typically performed using the known chemical reagent ions. For a reliable m/z calibration such known anchor ions are crucial. In many complex mixtures as for ambient aerosol, the signal observed at any nominal mass is composed of ions from numerous organic compounds and peaks are broad. Therefore, the exact mass of such a peak is unknown hindering an accurate m/z calibration.

In matrix assisted LDI-MS matrix molecules or matrix fragment ions have been used previously. However, the native matrix of our samples was highly variable, so that no ions provided suitable anchors for calibration. Therefore, we chose to spike our samples with AgNO₃. In the absence of such anchor ions an m/z calibration, even to integer accuracy, is not feasible.

Indeed, as mentioned in the reviewer's comment, we observed differences between the average spectrum containing AgNO₃ and the average spectrum without AgNO₃ which we related to changing ionization conditions. Hence, we did not report atmospheric data for AgNO₃-spiked samples. Rather, we used the AgNO₃-spiked spectra to obtain a first calibration (based on Ag⁺ anchor ions), used the first calibrate to define prominent, filter-specific anchor ions, and then calibrated the unspiked spectra using these anchor ions. This

procedure is explained in Section 2.2.2 Data treatment of the revised manuscript (unchanged from original) and was essential to perform an objective calibration. Because only prominent peaks were used in the second calibration step, the possibility of an influence of AgNO_3 on these samples is minimal.

Although we had mentioned these issues in the submitted manuscript, we have clarified them further following both reviewers' comments. We adapted the respective part of the manuscript to:

“... ”

The m/z calibration parameters were highly variable between different samples. Therefore, to determine unit mass resolution integration regions it was necessary to perform an m/z calibration on every single sample. However, unlike in the Aerodyne AMS (N_2^+ , O_2^+ , W^+), there are no dominant anchor ions present in these spectra that could be used for an m/z calibration. In absence of such ions, we performed a two-step calibration procedure. Each sample was spiked with silver nitrate (AgNO_3 , Sigma Aldrich, >99.8%) as an internal standard (approach illustrated in Fig. 1). In order to avoid the suppression of the sample signal, the internal standard (aqueous solution 500 ppt to 20 ppm) was only placed (as a droplet) on a small part of the sample. The 499 spectra from all positions on the measurement grid were separated and defined as (1) silver-spiked (lower panel in Fig. 1a), (2) silver-free (upper panel in Fig. 1a), and (3) intermediate-silver. Intermediate-silver cases were defined using the signal intensity in the regions of the mass spectrum where silver was expected, in comparison to adjacent silver-free regions of the spectrum, and were discarded. To calculate the first calibration, we calibrated the average silver-spiked spectrum of each filter sample, using the peaks of the silver monomer (m/z 107, 109 Th), dimer (m/z 214, 216, 218 Th), and trimer (m/z 321, 323, 325, 327 Th). We found that this calibration of silver-spiked was not directly applicable to the average silver-free mass spectrum. Possibly, spiking the filter region with aqueous AgNO_3 caused enough of a change to the surface of the filter sample to influence the ionization physics. This could affect the m/z calibration, since the delayed-pulse ion extraction in our instrument is not orthogonal to the ablation plume but nearly parallel. Therefore, a second calibration was obtained for the averaged silver-free spectra using prominent non-silver peaks present in both spectra. Such a two-step calibration is necessary to achieve accurate m/z calibrations of the silver-free spectra.

“... ”

The mass resolution as presented in Fig. 1b is computed based on the silver peaks since all other peaks are potentially influenced by a multitude of ions. Mass resolution and accuracy are independent properties; broad peaks can still be accurately calibrated in the absence of interference. We acknowledge that the mass accuracy in our dataset is poor. Therefore, we emphasize in the text that based on such accuracy an identification of contributing ions to a nominal mass is not possible (i.e. high-resolution analysis).

“... With this accuracy and resolution, neither distinguishing different ions at the same nominal mass nor estimating properties such as O/C and H/C for a nominal mass following Stark et al. (2016) was possible for the quartz-filter-LDI-MS measurements. Instead, the spectra were integrated to UMR sticks...”

Since the ions are not orthogonally extracted into the ToF unit, variability in the ionization process might be propagated into the ToF unit and, thereby, contribute to the poor accuracy of the m/z calibration. Since this variability was observed for the 1st calibration, which used only the Ag “anchor” ions which dominated the spectral intensity (so that a substantial interference by unidentified ions is very unlikely), it must be generally attributed to variability in the ionization and detection processes.

An observation of calibration inaccuracy in the calibration itself indicates that the calibration model was imprecise. The calibration model is the standard equation describing ion time-of-flight in the mass spectrometer, and would be violated if ions were at different distances from the detector or had different kinetic energy vectors, upon extraction. A non-uniform electric field in the spectrometer could also cause aberrances. These problems have been minimized in some instruments, but their minimization would reduce our ability to directly analyze filter samples, which is a major strength of our approach.

Page 5, Line 31. Remove “scales”

We removed this word.

Page 5, Line 34. It is the first time you mention “WSOC”, please mention the full name.

We added this information:

“... OC_{Sunset} was determined using the Sunset OC/EC analyzer and $(WSOM/WSOC)_{oAMS}$ through WSOM measurements (described in Section 2.1, WSOC being water-soluble organic carbon). ...”

Page5, Eq1. Scaling LDI intensity to AMS intensity may cause problems. In later analyses, we know that LDI detection has bias on different OA sources, but AMS detection should be rather equal. If the fraction of different OA sources changed, for example, LDI-insensitive source increased but LDI-sensitive source decreased in the same magnitude, this change will be clearly seen as a decrease in LDI data, but NOT in AMS data. Under such circumstances, the scaling factor has a dependency on source distribution. This might (NOT necessarily) affect the overall PMF results, for example the seasonal variation of PMF factors and the comparison of LDI-PMF and AMS-PMF. Thus, I think it is necessary to double check the results using unscaled LDI data.

The reviewer addresses one of the main challenges related to LDI-MS analyses: the quantification. During our experiments we discovered that (1) the signal recorded for the same sample during 1 day is fairly variable introducing a lot of variability and (2) that instrumental changes over time (possibly properties of the laser) introduced a decrease of signal with the measurement time. Consequently, the raw signal does not show a relation with OC, the sum of OC and EC or PM10 and is, therefore, not interpretable (Fig. S3, S4). Therefore, scaling to an external property is inevitable.

We updated the offline AMS source apportionment results used in the comparison (Fig 7) with the recently published data in Daellenbach et al. (2017). In the same time we also reworked Fig. 9 and 10 for an improved visibility. This resulted in small changes in Fig. 7 and the related text:

“ ...

As was the case for the comparison to NO_x and eBC_{tr} (Section 3.4.2), also in the comparison to the offline AMS traffic (HOA), a higher LDI-traffic/HOA was observed in summer which contributes to the low correlation coefficient ($R^2=0.04$). This might be related to the fact that HOA is only primary

without major contributions of traffic SOA or dust resuspension while LDI-traffic is potentially also influenced by aged/secondary traffic aerosol and resuspension. In contrast to summer, we observe a good agreement between LDI-traffic and HOA in winter. Thus even though a varying relative response factor (rRF) of the LDI-MS might contribute to these differences, these biases are not systematic but season dependent. As stated earlier, LDI-traffic is thought to be a mixture of primary tailpipe exhaust, aged/secondary tailpipe exhaust, and resuspended dust (as well as tyre break and engine wear). This will be further elucidated in Section 3.4.5.

LDI-BB was highly correlated with offline AMS-BBOA ($R^2=0.83$), yet the LDI-BB concentrations were higher than BBOA from oAMS, especially in northern Switzerland (with an LDI-BB:AMS-BBOA ratio between 1.4 and 4.1 for the different sites). A possible reason is also here the mixing of secondary components into LDI-BB when comparing to primary BBOA from oAMS. However, we cannot exclude that this effect is due to different rRFs of the LDI-MS for different compound classes.

The identified secondary components, LMW-OA ($R^2=0.45$) and bio-OA ($R^2=0.62$), also correlate with the corresponding OOA factors from the oAMS analysis, WOOA and SOOA, yet the correlation coefficients are smaller than for LDI-BB and BBOA. For these factors, differences between the two methodologies could be related to differences in the PMF performance or to differences in the response factors for different components in the LDI-MS (the LMW-OA: WOOA ratio is between 0.7 and 1.1 and the bio-OA: SOOA ratio between 0.2 and 0.4).

...”

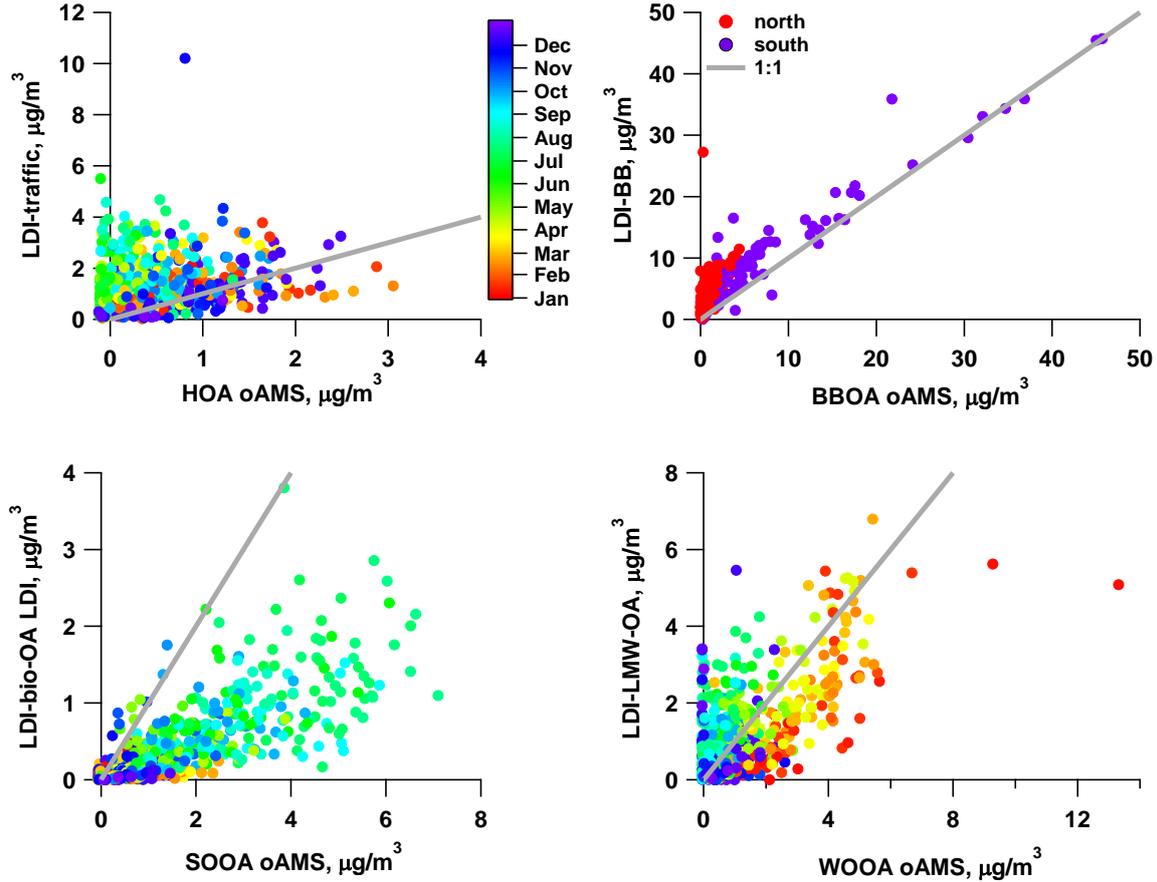


Figure 1: Comparison of LDI-MS to reference offline AMS source apportionment results for the sum of traffic related factors (LDI-traffic, $R^2_{LDI,oAMS}=0.04$), sum of wood burning related factors (LDI-BB, $R^2_{LDI,oAMS}=0.83$), bio-OA ($R^2_{LDI,oAMS}=0.62$), and LMW-OA ($R^2_{LDI,oAMS}=0.45$).

In response to the question, we added a paragraph on the analysis of relative response factors for LDI-MS in Section 3.4.3 of the manuscript:

“ ...

Overall, the source apportionment results based on LDI-MS data provide source separations with similar temporal behaviors as offline AMS and online AMS and ACSM analyses, yet seem to overestimate combustion related primary particle sources and underestimate secondary OA. However, we assume for this comparison that all factors give an equal response at a given concentration, i.e. the relative response factor (rRF_{LDI}) of all factors to be 1. The relative contribution of a factor k to the total signal observed with the LDI-MS ($r_{i,k,LDI}$) a specific measurement (i) depends on the factor concentration ($g_{i,k,LDI}$) as well as the sum of all factors ($\sum_k^p(g_{i,k,LDI})$) separated for the LDI-MS data. Assuming that differences between the LDI-PMF and the AMS-PMF arise solely from different rRF_{LDI} s for different factors, $r_{i,k,LDI}$ is a function of the AMS factor concentrations ($g_{i,k,AMS}$) and the rRF_{LDI} of this factor:

$$r_{i,k,LDI} = \frac{g_{i,k,LDI}}{\sum_k^p(g_{i,k,LDI})} = \frac{g_{i,k,AMS} * rRF_{k,LDI}}{\sum_k^p(g_{i,k,AMS} * rRF_{k,LDI})} \quad (4)$$

In order to determine rRF_{LDI} for the LDI-MS several strong assumptions are required: (1) the sum of traffic1 and traffic2 represents HOA, (2) the sum of WBeff, WBineff1 and WBineff2 represents BBOA, (3) LMW-OA represents WOOA, (4) bio-OA represents SOOA and (5) the AMS factors for which there is an LDI-MS equivalent, (1) - (4), are the only contributors to OA. In scenario 1 All the above assumptions are considered to be true. For scenario 1, we estimate rRF_{LDI-BB} as 1.84, $rRF_{LDI-LMW-OA}$ as 0.29 and $rRF_{LDI-bio-OA}$ as 0.40 (using LDI-traffic as reference factor, $rRF_{LDI-traffic} = 1.00$). The LDI-MS factor concentrations corrected using rRF_{LDI} show a close relation to the uncorrected LDI factor concentrations ($R^2_{LDI-traffic}=0.78$, $R^2_{LDI-BB}=0.94$, $R^2_{LDI-LMW-OA}=0.80$, $R^2_{LDI-bio-OA}=0.85$, Fig. 8).

We tested the sensitivity of the rRF_{LDI} estimates to the assumptions (1) and (2) being wrong (scenario 2) and only assumption (5) being wrong (scenario 3). In scenario 2, we alter scenario 1 by comparing the sum of traffic1 and bio-OA to SOOA and the sum of BBineff2 and LMW-OA to WOOA when computing rRF_{LDI} . For scenario 2, LDI-MS factor concentrations corrected using rRF_{LDI} show also a close relation to uncorrected LDI-MS factor concentrations ($R^2_{LDI-traffic2}=0.96$, $R^2_{BBeff+BBineff1}=1.00$, $R^2_{LDI-BBineff2+LMW-OA}=0.99$, $R^2_{traffic1+bio-OA}=0.96$). In scenario 3, we alter scenario 1 by considering also AMS factors without an equivalent in the LDI-MS PMF. Therefore, we compare LDI-traffic to the sum of HOA, COA, and SC-OA. Under these conditions LDI-MS factor concentrations corrected using rRF_{LDI} show a close relation to uncorrected LDI-MS factor concentrations ($R^2_{LDI-traffic}=0.75$, $R^2_{LDI-BB}=0.90$, $R^2_{LMW-OA}=0.74$, $R^2_{bio-OA}=0.85$). The differences in rRF_{LDI} between scenarios 1, 2, and 3 highlight the uncertainties caused by plausible violations of the underlying assumptions when determining rRF_{LDI} . Overall the estimates rRF_{LDI} are highly uncertain and their accurate determination needs to be the focus of future work. Given the high uncertainty of rRF_{LDI} and the good correlation between rRF_{LDI} corrected and uncorrected LDI-MS factor concentrations, we present uncorrected results without considering rRF_{LDI} .

Table 1: relative response factors (rRF) for LDI-MS analyses for 3 different scenarios.

scenario 1	scenario 2	scenario 3
$rRF_{LDI-traffic} = 1.00 \pm 0.00$	$rRF_{LDI-traffic2} = 1.00 \pm 0.00$	$rRF_{LDI-traffic} = 1.00 \pm 0.00$
$rRF_{LDI-BB} = 1.84 \pm 0.06$	$rRF_{LDI-BBeff+LDI-BBineff1} = 1.26 \pm 0.06$	$rRF_{LDI-BB} = 5.67 \pm 0.13$
$rRF_{LDI-LMW-OA} = 0.29 \pm 0.02$	$rRF_{LDI-LMW-OA+LDI-BBineff2} = 1.67 \pm 0.07$	$rRF_{LDI-LMW-OA} = 0.30 \pm 0.02$
$rRF_{LDI-bio-OA} = 0.14 \pm 0.01$	$rRF_{LDI-bio-OA+LDI-traffic1} = 0.39 \pm 0.02$	$rRF_{LDI-bio-OA} = 0.65 \pm 0.02$

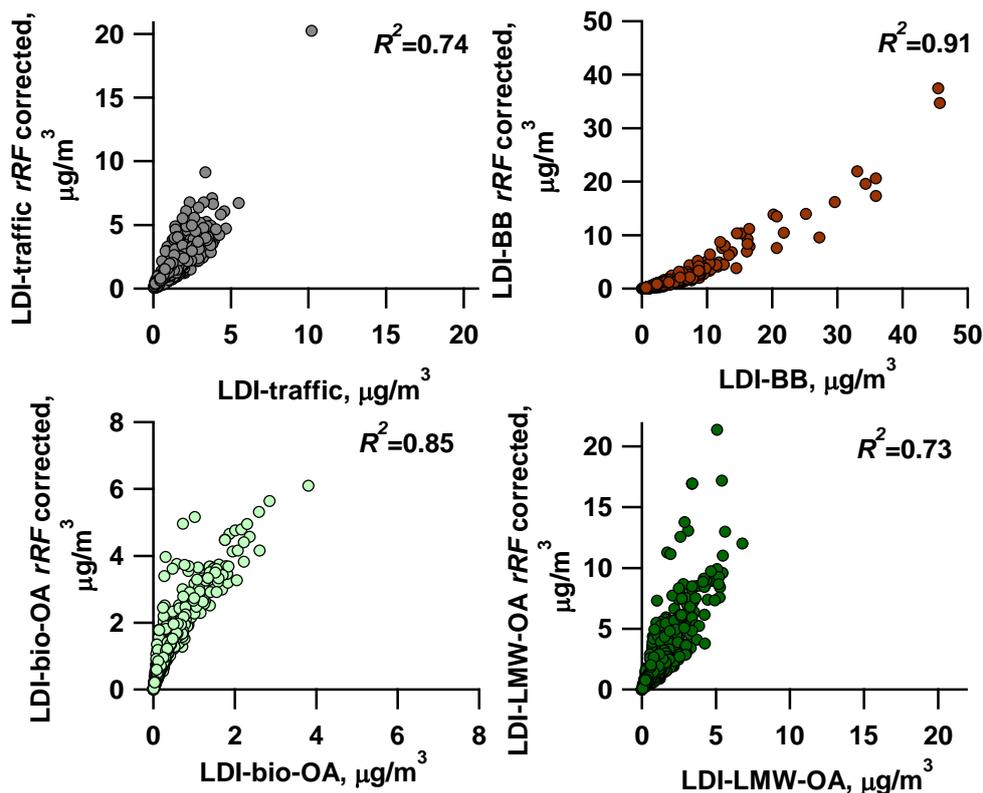


Figure 2: Comparison of LDI-MS factor concentrations corrected using relative response factors (*rRF*) to uncorrected LDI factor concentrations (Scenario 1).

...”

Additionally, we also add a comment in the conclusion:

“... Other components are more difficult to compare quantitatively because of different source separations in PMF as well as differences in the relative response factors (*rRF*) of OA components. *rRF* determined in this study are uncertain and, therefore, not used for correcting the LDI-PMF results. ...”

Page 7, Line 22-25. The fitting parameter in November is clearly different (Fig.2 red dots), though maybe not significantly.

The purpose of these exercises was to obtain an error model as presented in Section 3.1 and in the Supplement. There are only three repeats. Therefore, it is hard to judge whether the experiment in question is an outlier or the result of a broad distribution. We report the resulting uncertainty of the error model in the manuscript as following:

“... The relative error was related to the absolute signal intensity for each *m/z* for each sample and experiment (Fig. 2), with an asymptotic value of only $\sim 9 \pm 7\%$ at high signals (average and standard deviation of fitting Eq. 3). The average absolute error for small signals was 102 ± 48 a.u. (from Eq. 3)...”

Page 8, Line 33. “: : of which 84, 94, 120, and 177 correlated with NO_x...”. 84amu is not marked in Fig.4a, and it does not show a strong correlation with NO_x (R ~0.2 as I read from the figure)

The sentence has been corrected:

“... The m/z 84 94, 101, 120, 143, 165, and 177 Th, also detected in tunnel samples, showed high relative intensities in Bern in summer, of which 94, 120, and 177 Th correlated with NO_x, which suggests a high contribution of traffic emissions at this location, consistent with previous observations (Zotter et al., 2014). ...”

Page 9, Line 15. “Both traffic1 and traffic2 showed a relationship to NO_x, only traffic2 correlated with eBC_{tr} (Table S1)”. What does the “relationship” mean? The correlation between traffic 1 and NO_x is very weak. From Table S1, I think the traffic 1 has neither correlation with NO_x, nor with eBC; while traffic 2 may correlate with both.

We rephrased the respective part of the manuscript:

“... Both traffic1 and traffic2 showed increasing concentrations with increasing NO_x levels. However traffic1/NO_x and traffic2/NO_x were seasonally variable ($R_{p,traffic1,NO_x}=-0.17$, $R_{p,traffic2,NO_x}=0.24$, Fig. 6, Fig. S9). While traffic2 correlated with eBC_{tr}, traffic1 did not show such a dependency (Tab. S1, Fig. 6, Fig. S9). ...”

Page 9, Line 16-17. “This suggests that traffic2 is related to primary emissions from the combustion process and that traffic2 is also influenced by other processes, like e.g. secondary production.” For people who are not familiar with traffic tracers, e.g. NO_x and eBC, this interpretation is hard to follow. It is good to explain this a bit more.

We rephrased the respective part of the manuscript:

“... The correlation with eBC_{tr} suggests that traffic2 is related to primary emissions from the combustion process. Based on the lack of correlation between traffic1 and eBC_{tr}, traffic1 might also be influenced by other processes, such as e.g. secondary production. ...”

Page 10, Line 31-33. “The contribution of the last remaining factor showed an exponential increase with temperature, similar to terpene emissions and biogenic SOA (Leaith et al., 2011), suggesting this factor to be strongly influenced by biogenic SOA production. Therefore, this factor was termed bio-OA.” Some other atmospheric processes, other than VOC emissions, are temperature-dependent, such as VOC oxidation and photochemistry. Is there additional evidence to validate the bio-OA factor, such as higher fractional contribution in rural sites than in urban sites?

Using yearly average NO_x concentrations as a metric for rurality/urbanity, Payerne sticks out as rural (low NO_x) and Bern as urban (high NO_x). In Payerne bio-OA has a higher relative contribution (11%) than in Bern (3%) which are the highest and lowest in the entire dataset (see Table 2 in manuscript). Qualitatively, the same was observed in the AMS data (38% SOOA in Payerne and 21%; only S. Vittore showed a lower contribution at the expense of higher BBOA, Daellenbach et al., 2017).

In order to make this information more visible we added yearly NO_x concentrations to Tab. 2 and sorted the sites according to the yearly average NO_x concentration.

Such information was added to the manuscript:

“...Bio-OA had a highest relative contribution at the most rural site in the dataset (Payerne: yearly average bio-OA 11%, yearly average NO_x concentration: 9 ppb) and the lowest at the most trafficated site (Zurich: yearly average bio-OA 3%, yearly average NO_x concentration: 48 ppb). ...”

References:

Daellenbach, K. R., Stefenelli, G., Bozzetti, C., Vlachou, A., Fermo, P., Gonzalez, R., Piazzalunga, A., Colombi, C., Canonaco, F., Hueglin, C., Kasper-Giebl, A., Jaffrezo, J.-L., Bianchi, F., Slowik, J. G., Baltensperger, U., El-Haddad, I., and Prévôt, A. S. H.: Long-term chemical analysis and organic aerosol source apportionment at nine sites in central Europe: source identification and uncertainty assessment, *Atmos. Chem. Phys.*, 17, 13265-13282, <https://doi.org/10.5194/acp-17-13265-2017>, 2017.