Response to reviewer comments for manuscript: "On the calibration of FIGAERO-ToF-CIMS: importance and impact of calibrant delivery for the particle phase calibration"

Ylisirniö et al.,

We thank the reviewer for his/her constructive comments regarding our paper. Below we will address the specific issues point by point. The reviewer's comments are in black and our answers are in blue. Changes to the Manuscript or Supplement Information are highlighted in red.
Line numbers before the red response text refer to line numbers in the modified manuscript.

Additional changes by authors:

We added additional discussion about potential effect of collected aerosol mass loading to Section 3.3.

Line 281:

"An additional aspect that has been reported to shift T_{max} values is the amount of collected aerosol mass on the PTFE filter (Huang et al., 2018), becoming important when collected particulate mass is around several micrograms. We tested the mass loading effect by collecting different amounts of atomized PEG's up to 200 ng of mass and found no clear difference between measured T_{max} values (data not shown). However, as collected aerosol mass on the FIGAERO filter can easily reach microgram amounts, especially when sampling in highly polluted environments and as we did not rigorously test how T_{max} values behave above 200 ng, we suggest that this effect is investigated further in future publications."

Reviewer 1:

General comments:

1) Dependence on the solution concentration. The authors demonstrate that by adjusting the diameter in the evaporation model, they can reproduce the results for different solution concentrations. In order to have this in line with the subsequent reasoning about the SEM image, they should show different SEM images with different solution concentrations showing that indeed smaller structures are deposited on the filter in the case of lower solution concentrations. If the diameter really controls this behavior, also a calibration with the atomizer and particles around 1 µm could support this.

The reviewer makes a good point regarding the additional SEM figures with different solution concentrations. Below is a SEM figure showing deposition using a solution of 0.25 g L^{-1} which is a much higher concentration than the 0.01 g L^{-1} used in Fig. 5b) of the manuscript. Careful analysis

may reveal some differences between the “ring structures” in these two cases. But, in our opinion, it is almost impossible to do a quantitative analysis of the true size and mass of the evaporating unit from the filter based solely on this type of SEM pictures. While it is easy to measure the diameter and width of the “ring”, we have no information about the thickness of the layer. Additionally, part of the deposited compound/mass may be hiding inside the filter and cannot be seen with this method.

Thus, it is our understanding that a more thorough analysis using a higher number of different solution concentrations would not lead to more relevant information. We will continue to just use the SEM pictures to qualitatively verify that the two deposition methods (syringe vs atomization) lead to very different structures of the deposited material on the filter which impact the evaporation behaviour. However, we will more strongly emphasise the qualitative nature of the SEM pictures in the modified manuscript.

Line 242:

“We want to emphasize that as SEM cannot distinguish deposited material situated inside the filter or measure the layer depth, the images shown here should be considered only as qualitative evidence.”

Unfortunately, creating monodisperse particles in the range of 1 µm is not trivial with our equipment. Optimising the particle generation and size selection for that size range was outside of the scope of this study.

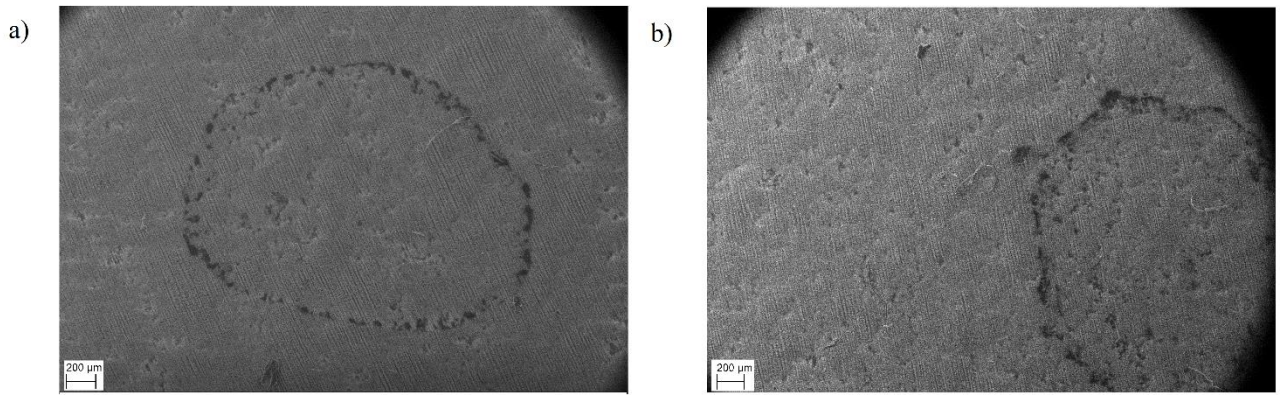

Figure 1. Panel a) Shows Fig.5 b) from manuscript showing 3 µl of PEG-8 with concentration of 0.01 g L^{-1}. Panel b) shows SEM figure of 3 µl of PEG-8 with concentration of 0.25 g/l in ACN deposited on the filter.

2) The authors base their reasoning mainly on the SEM images of the FIGAERO filter for the atomization method and the syringe method. However, they only show SEM pictures for one substance. While I don’t doubt their conclusion, other SEM images should be added. The authors make the statement that the vacuum in the SEM could evaporate all the other substances than PEG-8, but how can they than conclude that it is not the SEM sample preparation (i.e. bringing the filter into a vacuum), which causes the structures observed on the filter?

We argue that the formation of the ring structure is caused by the evaporation of the initially deposited droplet, i.e., the evaporation of the solvent, ACN. Changing the rate of evaporation (i.e., by evaporating at ambient or SEM pressure) may change the exact size and shape of the “ring

structure", but the structure will still be formed and be very different from the same mass deposited as sub-micron particles.

The only difference between normal FIGAERO sample treatment and SEM sample treatment is that SEM is operated in vacuum while in the FIGAERO the samples are not exposed to lowered pressures. PEG-8 was selected for the screening due to its low vapour pressures, to minimise the likelihood of any evaporation of the example calibration compound in the vacuum of the SEM which might skew the results from the SEM. While it is possible to use some solid compound like citric acid, which would not easily evaporate from the filter even in vacuum, getting these SEM pictures would be a considerable effort, especially under the current circumstances with limited personnel available due to the pandemic situation.

3) Impact of using different calibration methods. When showing the different VBS systems, I would like to see also a comparison to a VBS derived using a group contribution method or a fit to it (as e.g. in Stolzenburg et al., 2018 or Mohr et al., 2019). This would indicate which calibration method is more in line with this widely used approach, which does not rely on a direct volatility measurement.

Determining the VBS distributions with group contribution methods and comparing them to VBS distributions determined using direct volatility measurements is indeed an interesting topic, but we decided to leave that analysis out from our manuscript as we think it to be a whole topic of its own. One important issue is that most of the parameterisations were developed for measurements of gaseous compounds. When a thermal desorption step is included, the extend of thermal decomposition must be considered, i.e. that low-volatile but thermally labile compounds are detected as small fragments which will lead to a significant overestimation of their C^* values (see e.g. Buchholz et al 2020, Lopez-Hilfiker et al. 2015, Schobesberger et al. 2018, Stark et al. 2017).

However, we are currently working on this topic and will discuss it in more detail in future publications.

Minor:

1) P.1, l.31-32: I am missing a short discussion on other volatility measurement techniques, e.g. VTDMA setups. Please add this here.

Added reference to VTDMA.

2) P.2, l.39: I am missing some laboratory studies from the CLOUD team published recently, e.g. Wang et al. (2020, Env. Sci. Techn. & Nature), Stolzenburg et al. (2018, PNAS). Also missing is Mohr et al. (2019, Nat. Commun.). In all these studies the FIGAERO-CIMS was deployed quite successfully and they could be mentioned here for completeness.

Added Stolzenburg et al. 2018 and Mohr et al. 2019 to references listed at this line and additionally also added Wang et al. 2020 to the reference list of published calibration lines.

3) P.2, l.49: Also Wang et al. (2020, Env. Sci. Techn.).

Added Wang et al. 2020 to reference list, and updated Figure 1 and Figure S2 with their calibration line.

4) P.5, l.144: Did you constrain the width of the lognorm fit for the desorption? This could be necessary especially for unknown compounds, which might have isomers or fragments on the same mass yielding a bimodal structure.

The reviewer is right that often the fit needs to be constrained. As the calibration compound thermograms are “ideal”, lognormal fit can be applied to the whole thermogram, but in “real” data the fit usually needs to be constrained around the peak of the thermogram, especially if the thermogram is bimodal or broadened by the presence of isomers and/or fragments from thermal decomposition. Note that in the case of the presence of multiple compounds of different volatility, the T_{max} value may represent only the volatility of the dominant compound and ignores the contribution of the minor compounds with that sum formula. But it is also possible that the T_{max} value represents an “average” over multiple compounds, especially if the thermogram peaks of the isomers/fragments are too close to be distinguished. For such cases, the T_{max} method cannot capture all of the volatility information and a more sophisticated method is needed to separate the compounds (e.g. Positive Matrix factorisation, Buchholz et al. 2020)

5) P.6, l.185-187: If the inlet is initially at a different temperature, the supply of a constant heat rate will yield a different thermogram, as it takes longer to achieve the corresponding temperatures allowing more time for evaporation. Is this considered in then model? And how can we use calibrations performed at one temperature in comparison to measurements at different temperatures? Could the model resolve this?

The FIGAERO uses filters made of only PTFE. The filter holder and the moving tray are also machined from PTFE. One reason for that choice is the material's chemical inertness. But the main reason for choosing PTFE is its low thermal conductivity. Consequently, the temperature of the deposit is more directly controlled by the heat of the N2 flowing through the filter, which is measured immediately upstream and hence well understood. The model therefore does not explicitly consider the heating of the filter material. But it does allow for "non-ideal" heating of certain parts of the deposit. The description of that non-ideality is fairly crude (details in Schobesberger et al., ACP, 2018). It has not been modelled as a function of heating rate, as that has not appeared necessary. The main "job" of the non-ideal heating in the model is to produce thermogram tails; it hardly affects T_{max}.

If the reviewer's latter two questions refer to the *ambient* temperature of the calibrations, the model is not currently set up to resolve resulting issues.

If the reviewer refers to extending the experimentally obtained "calibration curve" to T_{max} values beyond those provided by the observed T_{max} (illustrated e.g. in Fig. S2, and subject of Fig. S3), the model could in principle do that, but it would require asserting a relationship between saturation vapor pressures and vaporization enthalpies (e.g., see Fig. 7 in Schobesberger et al., 2018).

6) P.7, l.202: Repeat the atomizer solution concentration to put it into the context with the syringe concentrations. Also mention here the mode diameter of the particles used for calibration or even calculate the deposited mass for this type of calibration compared to the syringe method. This would put the two methods into comparison here.

Added more information to the section.

Line 208 onwards:

"For comparison, the starting concentration of the atomizer solution was 0.5 g L-1 for each compound. The solution concentration gradually increased as the solvent evaporated from the solution. This led to a polydisperse, log-normal-shaped aerosol population with a mode diameter of 50 nm. From this distribution, particles equivalent to ~200 ng of aerosol mass were sampled onto the FIGAERO filter before desorption."

7) P.8, l.231: Instead of mentioning the different scale, I would like to see a fourth panel in Fig 4 showing the filter in the same scale as in Fig. 5c! This would help to directly compare the different structures deposed on the filter.

Added fourth panel to Figure 5 (panel c) with 10 μm scale, taken as same sample as Fig. 5a). Original Fig. 5c) is now Fig 5d).

8) P.8, l.240: Also the larger diameters needed to explain the syringe calibration with model point into that direction. This is an important supportive argument and should be mentioned here.

We added the following to the revised manuscript:

Line 254 onwards:

Indeed, it was by building on these assumptions that the evaporation model succeeded in reproducing the observations in Fig. 4. With the much smaller surface area of the syringe deposited material, it requires more time to evaporate all the PEG-8 than from the equivalent amount of deposited aerosol particles. This time delay directly translates to a shift to higher observed T_{max} values. The desorption model mimics this change in surface-to-volume ratio by increasing the initial size of the modelled evaporating particle to 1.3 μm and 11 μm. But note that there are no individual spherical particles of that size on the filter.

9) P.8, l.251: Any hints why the different inlet behaves that way?

The positioning of the filter thermocouple affects the measured temperature of the desorption flow. It also affects the exact offset between the measured temperature and the temperature at the filter surface. To position the thermocouples in exactly the same position in two different inlets is quite difficult, so it is relatively hard to achieve exactly comparable measurements using separate inlets.

This is one of the reasons why a temperature calibration is necessary for each FIGAERO inlet and also every time the inlet is disassembled as disassembling the inlet may affect the position of the thermocouple.

10) P.8, l.254: Why does it fail for PEG-8? Please elaborate on that.

In the model simulation, molecules that have desorbed from deposited particles are subsequently interacting with instrument surfaces, which are experiencing the same temperature ramp as the deposit, before being measured. That interaction is simulated by 100% initial absorption, and desorption as a function of $C^*(T)$ and an optional instrument constant. (The latter is indeed the main tuning parameter for the model to reproduce observed T_{max} values). These interactions cause a delay. The delay translates to higher T_{max} and increases with decreasing C^* as well as decreasing particle size. Consequently, differences in T_{max} due to particle size disappear when C^* and particle size are sufficiently small, as T_{max} becomes controlled by those vapor-surface interactions. This is what we observe in the model outputs for PEG-8 (and somewhat also for PEG-7).

We believe that that is a shortcoming of the model, rather than of the experiments, because: (a) the thermograms were experimentally very well reproducible, and (b) the thermograms (in particular for PEG-8) were narrower for 80-nm particles than for 300-nm particles. From the model's point of view, observation (b) would suggest more ideal heating in the 80-nm case than in the 300-nm case, while the lower T_{max} would suggest less vapor-surface interactions. To affect that, 80-nm particles would need to be deposited somehow substantially differently on the filter than 300-nm particles. More likely instead, the model's current treatment of both non-ideality of heating and vapor-surface interactions (Schobesberger et al., 2018) are insufficiently close to reality in this case.

11) P.10, l.295: Move "A more detailed description of the SOA production is shown in Ylisirniö et al., 2020." in front of the preceding sentence.

Sentence moved.

12) P.10, l.299: 200 g mol-1 seems quite low for alpha-pinene HOMs, e.g. Tröstl et al. used 300 amu as mean mass.

The reviewer is correct that 200 g mol^{-1} is small compared to the mean molecular mass of alpha-pinene HOMs. However, when C^* is plotted in logarithmic space as in Fig.9, the change from 200 g mol^{-1} to 300 g mol^{-1} becomes negligible, as can be seen in the figure below. We therefore think that 200 g mol^{-1} is adequate for our purposes.

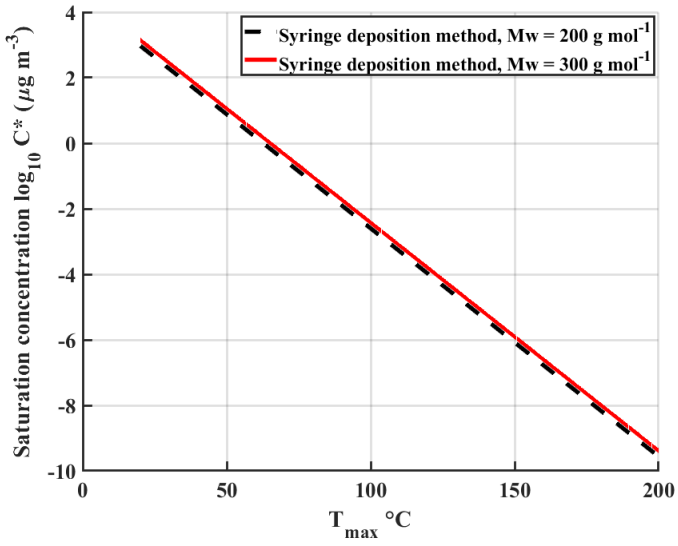

Figure 2. Figure 9 panel d) syringe deposition method theoretical line calculated with 200 g mol^{-1} and 300 g mol^{-1}.

13) P.11, l.355: Seems logical, but extremely difficult to realize in the lab. What would be the best alternative?

We are not completely sure what the reviewer means with this question as the sentence in the line 355 reads: "We note that these P_{sat} values have not been verified by other studies and are subject to corrections, but want to point out that harmonizing further FIGAERO calibrations by using PEGs would make future FIGAERO measurements more comparable to each other."

We want to note that by word "harmonizing" we don't mean a rigorous ISO-standard style calibration procedure, but simply that each FIGAERO would be calibrated with same compounds and using same P_{sat} values.