Reply to RC 1

Manuscript information:

- Title: Non-target analysis using a gas chromatography with time-of-flight mass spectrometry: application to time series of fourth generation synthetic halocarbons at Taunus Observatory (Germany).
- Authors: Fides Lefrancois, Markus Jesswein, Markus Thoma, Andreas Engel, Kieran Stanley, and Tanja Schuck
- MS No.: amt-2020-488
- MS type: Research article
- Iteration: Final response (AMT Discussions)

We would like to thank the reviewer for the constructive and detailed comments.

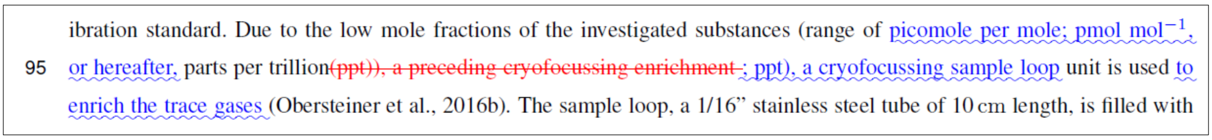
Point-by-Point reply:

1. *“The analysis raised a question for me (but that doesn’t need to be addressed in the manuscript). The question is why don’t all compounds work equally well as internal standards? And if they don’t all work well, what is the uncertainty of the non-target compounds for any specified period? They might behave well or might not, it seems to me. Also, given the stated precisions of the calibration standards, it is then surprising that the correlations between standard responses show percent errors in the 10% range. I’m not sure how to interpret that.“*

We do not exactly know what effects the different substances and why they do not behave the same way in different standards or measurements. Some substances could be influenced by changing H_2O-amounts or different ionization processes. We investigated several parameters of the compounds (e.g. retention times, signal intensities, compound similarities) but we could not find any conclusion yet. The internal calibration standard precision does not have an influence on this approach. Finally, we have to say, it is also a system dependent issue. We tried to point this more in section 3.2.1 *Relative Response Factor* and answer on point 5.

2. *L 81. Suggest eliminate “preceding” or change to “…(ppt)), sample trace gases are enriched by cryofocussing in a sample loop.”*

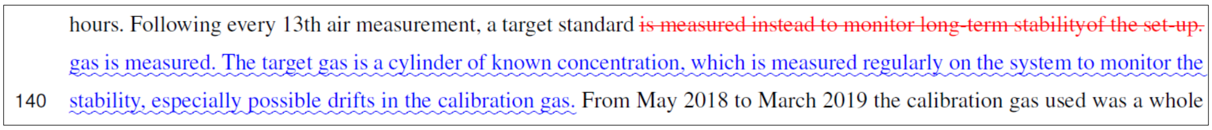
We changed it to: “(…), a cryofocussing sample loop unit is used to enrich the trace gases (Obersteiner et al., 2016b).”



ibration standard. Due to the low mole fractions of the investigated substances (range of picomole per mole; pmol mol^{-1}, or hereafter, parts per trillion~~(ppt)), a preceding cryofocussing enrichment~~ ; ppt), a cryofocussing sample loop unit is used to enrich the trace gases (Obersteiner et al., 2016b). The sample loop, a 1/16” stainless steel tube of 10 cm length, is filled with

3. *L113. Though I think I know what you mean, could you better describe what a "target" standard is?*

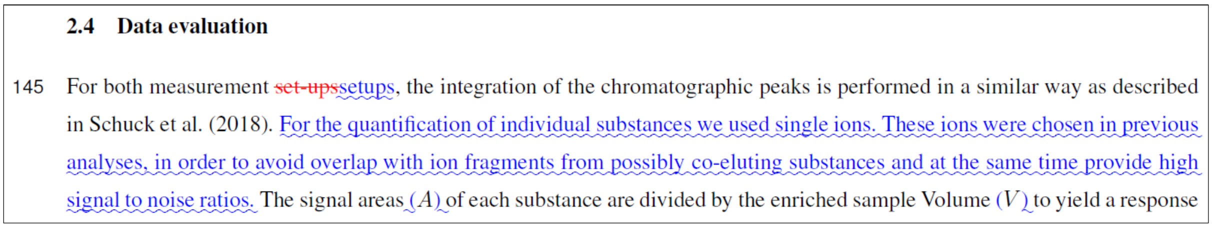
We added a comment on that:



hours. Following every 13th air measurement, a target standard ~~is measured instead to monitor long-term stabilityof the set-up.~~ gas is measured. The target gas is a cylinder of known concentration, which is measured regularly on the system to monitor the stability, especially possible drifts in the calibration gas. From May 2018 to March 2019 the calibration gas used was a whole

4. *L120. It would be helpful to describe some more detail of the method. For example, I couldn't find anywhere how the quantitation was accomplished with the QTOF data. Was a single selected ion used for each compound, or the sum of major fragments, or some integration of a peak that has been deconvolved from the total ion chromatogram? Were different methods of sample integration tried? Perhaps there is also a relationship between mass and total ion current that could be used to quantitate certain classes of compounds (at least within 25%)? Other questions: mass resolution of TOF?*

We have followed the reviewer's suggestion and added some more detail on the quantification to the manuscript:



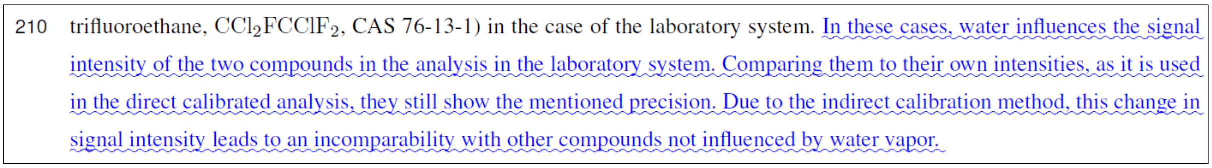
2.4 Data evaluation

For both measurement ~~set-ups~~setups, the integration of the chromatographic peaks is performed in a similar way as described in Schuck et al. (2018). For the quantification of individual substances we used single ions. These ions were chosen in previous analyses, in order to avoid overlap with ion fragments from possibly co-eluting substances and at the same time provide high signal to noise ratios. The signal areas (A) of each substance are divided by the enriched sample Volume (V) to yield a response

The mass resolutions of the used TOFMS systems are given in the quoted manuscripts in section 2 *measurements*.

5. *L177. Although HCFC_141b elutes near water, it shows excellent precision. So not sure why this might be excluded. Or it might be interesting to learn how water vapor might influence the results.*

We added a comment on that:



trifluoroethane, CCl_2FCClF_2, CAS 76-13-1) in the case of the laboratory system. In these cases, water influences the signal intensity of the two compounds in the analysis in the laboratory system. Comparing them to their own intensities, as it is used in the direct calibrated analysis, they still show the mentioned precision. Due to the indirect calibration method, this change in signal intensity leads to an incomparability with other compounds not influenced by water vapor.

6. *L179. Not sure if the plots are artificial data from some "arbitrary substances" or if the actual compounds are just not named here. Could you clarify?*

The figure is a „dummy"-data-set, to explain the methodology schematically. We added a clarification in the text.

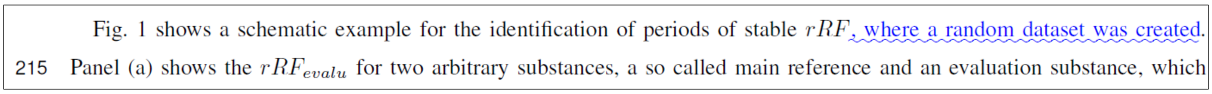


Fig. 1 shows a schematic example for the identification of periods of stable rRF, where a random dataset was created. Panel (a) shows the rRF_{evalu} for two arbitrary substances, a so called main reference and an evaluation substance, which

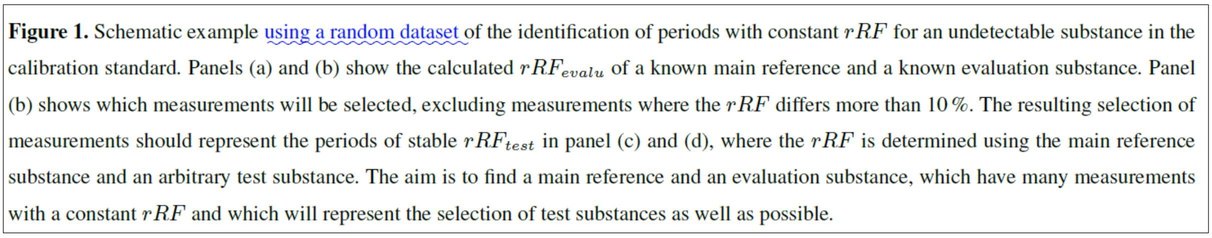
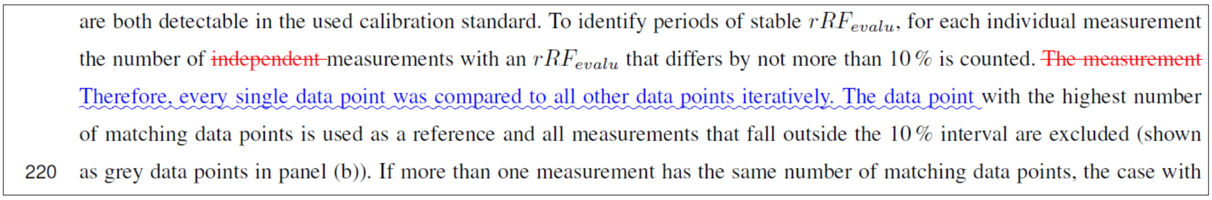


Figure 1. Schematic example using a random dataset of the identification of periods with constant rRF for an undetectable substance in the calibration standard. Panels (a) and (b) show the calculated rRF_{evalu} of a known main reference and a known evaluation substance. Panel (b) shows which measurements will be selected, excluding measurements where the rRF differs more than 10 %. The resulting selection of measurements should represent the periods of stable rRF_{test} in panel (c) and (d), where the rRF is determined using the main reference substance and an arbitrary test substance. The aim is to find a main reference and an evaluation substance, which have many measurements with a constant rRF and which will represent the selection of test substances as well as possible.

7. *L182. It is not clear how the 10% criterion for rejection is applied. Is this from point to point, or relative to some average?*

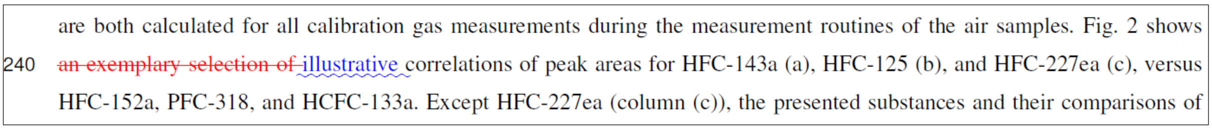
We clarified this by changing the manuscript as follows:



are both detectable in the used calibration standard. To identify periods of stable rRF_{evalu}, for each individual measurement the number of ~~independent~~ measurements with an rRF_{evalu} that differs by not more than 10 % is counted. ~~The measurement~~ Therefore, every single data point was compared to all other data points iteratively. The data point with the highest number of matching data points is used as a reference and all measurements that fall outside the 10 % interval are excluded (shown as grey data points in panel (b)). If more than one measurement has the same number of matching data points, the case with

8. *L197. Not sure if you mean "exemplary", as in "best example of the group", or are these just examples of some of the compounds. (also in Figure caption).*

These are illustrative examples of some of the investigated substances to demonstrate that some substances correlate well, some not. We clarified this as follows:



are both calculated for all calibration gas measurements during the measurement routines of the air samples. Fig. 2 shows ~~an exemplary selection of~~ illustrative correlations of peak areas for HFC-143a (a), HFC-125 (b), and HFC-227ea (c), versus HFC-152a, PFC-318, and HCFC-133a. Except HFC-227ea (column (c)), the presented substances and their comparisons of

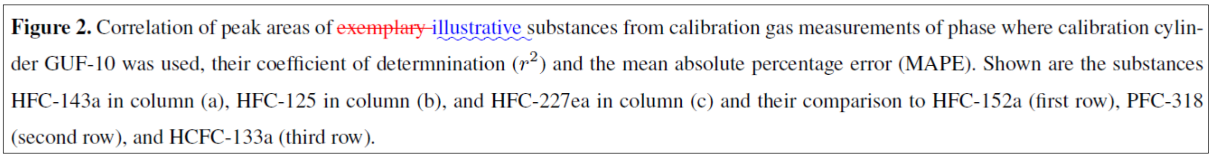
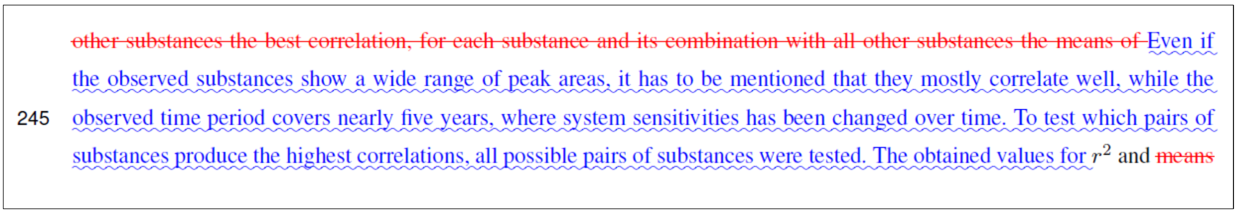


Figure 2. Correlation of peak areas of ~~exemplary~~ illustrative substances from calibration gas measurements of phase where calibration cylinder GUF-10 was used, their coefficient of determnination (r^2) and the mean absolute percentage error (MAPE). Shown are the substances HFC-143a in column (a), HFC-125 in column (b), and HFC-227ea in column (c) and their comparison to HFC-152a (first row), PFC-318 (second row), and HCFC-133a (third row).

9. *Figure 2. As I understand it, this figure compares the peak areas of compound pairs in the same calibration standard over the time of the study. Could you comment on the very large range of peak areas that were observed? Is this a characteristic of the TOF?*

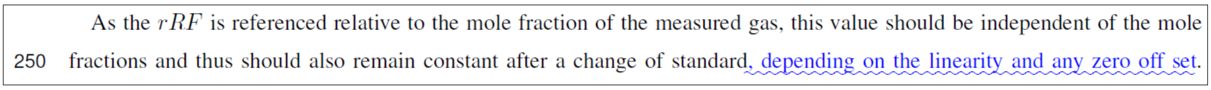
We do not believe that this is a special characteristic of the TOF. It should be noted that measurements cover a period of nearly five years, where the sensitivity can change due to e.g. degradation of filaments and detectors, tuning of the instrument etc. We have added a comment on that in the manuscript:



~~other substances the best correlation, for each substance and its combination with all other substances the means of~~ Even if the observed substances show a wide range of peak areas, it has to be mentioned that they mostly correlate well, while the observed time period covers nearly five years, where system sensitivities has been changed over time. To test which pairs of substances produce the highest correlations, all possible pairs of substances were tested. The obtained values for r^2 and ~~means~~

10. *L204. Note that the independence of rRF will also depend on linearity and any zero offset.*

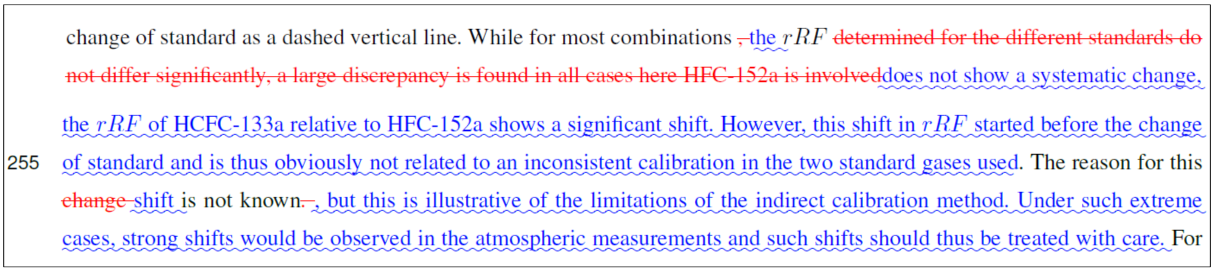
This is true, we mention this now in the manuscript:



As the rRF is referenced relative to the mole fraction of the measured gas, this value should be independent of the mole fractions and thus should also remain constant after a change of standard, depending on the linearity and any zero off set.

11. *L209. The observed shift of 152a relative to 133a deserves some comment. Presumably there was no similar shift in the time series ambient measurements of either gas. So, this behavior, though maybe rare, would seem to be a major red flag in applying the proposed method with confidence.*

The reviewer is correct in pointing out that this is a critical point. We added a note of caution explaining that in some cases such outliers may occur, which may not be caught by the preprocessed data analysis and its filtering method.



change of standard as a dashed vertical line. While for most combinations ~~,~~the rRF ~~determined for the different standards do not differ significantly, a large discrepancy is found in all cases here HFC-152a is involved~~does not show a systematic change, the rRF of HCFC-133a relative to HFC-152a shows a significant shift. However, this shift in rRF started before the change of standard and is thus obviously not related to an inconsistent calibration in the two standard gases used. The reason for this ~~change~~ shift is not known~~.~~, but this is illustrative of the limitations of the indirect calibration method. Under such extreme cases, strong shifts would be observed in the atmospheric measurements and such shifts should thus be treated with care. For

12. *L213. There are a number of compounds that have drifts or anomalies that prevent them from being used as "reference" compounds. Does this have any implication on how these are used for direct calibration? Do these standards cause the sample mixing ratios to be flagged? The authors also suggest that there are a number of potential factors that will influence the relative responses. In cases of outliers or large shifts (such as 152a), have the authors determined specific causes for the deviations?*

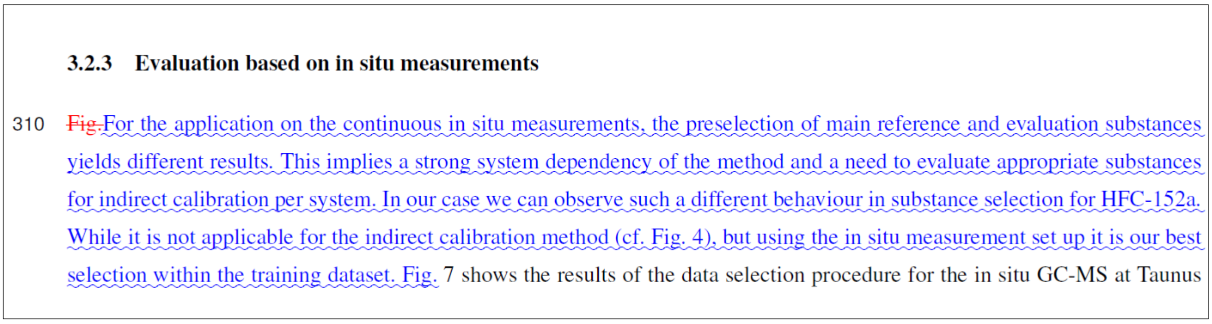
The drift in relative sensitivities between two compounds does not have any direct implication for the direct calibration. So, there is no need to flag these substances in the direct calibration. As mentioned above, we investigated several potential reasons for the deviations (differences between the retention times, different signal to noises/signal intensities) of this occurrence. But – using these preselected substances – we could not find a regularity. This could be an interesting investigation in further studies, but we prefer not to elaborate on this in the present study.

13. *L214. Not sure of meaning…change "suited" to "suitable"?*

Changed that.

14. *L251. I was interested to see that 152a was selected as a reference standard for the in-situ measurement evaluation, though there was a problem with this compound in the canister analysis. As noted, this is disturbing and deserves comment.*

We added a comment on that:



3.2.3 Evaluation based on in situ measurements

~~Fig.~~For the application on the continuous in situ measurements, the preselection of main reference and evaluation substances yields different results. This implies a strong system dependency of the method and a need to evaluate appropriate substances for indirect calibration per system. In our case we can observe such a different behaviour in substance selection for HFC-152a. While it is not applicable for the indirect calibration method (cf. Fig. 4), but using the in situ measurement set up it is our best selection within the training dataset. Fig. 7 shows the results of the data selection procedure for the in situ GC-MS at Taunus