

Reply to RC 4

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 Anja Claude for the constructive and detailed comments.

Point-by-Point reply:

1. *Close work with European metric institutions has brought up a discussion on “correct vocabulary”. Among this the terminus “mole fraction”. As “mole” is a unit, NMIs requested the use of “amount fraction” instead und the correct unit would be “nmol/mol” (instead of ppt) - just a remark, as I came across this discussion often recently.*

We agree that the correct unit would be pmol/mol. As the standards we use are prepared gravimetrically, we think that the wording “mole fraction” describes the measured quantities more precise than “amount fraction”. For simplicity we will use ppt rather than pmol/mol and have added an explanatory sentence to the manuscript:

ibration standard. Due to the low mole fractions of the investigated substances (range of [picomole per mole; pmol mol⁻¹](#)
95 [or hereafter](#), parts per trillion([ppt](#))), ~~a preceding cryofocussing enrichment~~; [ppt](#)), a cryofocussing [sample loop](#) unit is used to

2. *Introduction: l. 47, “These hydro(chloro-)fluoroolefines are the so-called fourth generation of synthetic halocarbons...” - there are no other fourth generation synthetic halocarbons?*

We changed the sentence as follows:

50 HFC-1234ze(E), CAS 29118-24-9), and HCFO-1233zd(E) (E-1-chloro-3,3,3-trifluoroprop-1-ene, ~~HCFC-1233zd(E)~~[trans-CF₃\(](#)
[HCFC-1233zd\(E\)](#), CAS 102687-65-0). ~~These~~[In the following we will use the H\(C\)FO nomenclature for the hydro\(chloro-](#)
[\)fluoroolefines](#)~~are~~[, as the HFC-nomenclature is not made for compounds with a double bond. These H\(C\)FOs are examples of](#)
the so-called fourth generation of synthetic halocarbons. Due to their carbon double bond, ~~HFO and HCFO~~[HFOs and HCFOs](#)

3. l.51: “However, some HFO, as some HFC and HCFC, can form the very persistent and toxic trifluoroacetic acid (TFA) as...” check commas and the use of "some" “ Are the three substances of this paper among the TFA forming HFOs?

We clarified this as follows:

have an ODP of zero (Patten and Wuebbles, 2010; Orkin et al., 2014). However, some ~~HFO, as some HFC and HCFC~~HFOs, e.g. the HFO-1234yf, and also some HFCs and HCFCs, can form the very persistent ~~and toxic~~-trifluoroacetic acid (TFA), as the main breakdown product in the atmosphere (Burkholder et al., 2015). TFA ~~is known to cause negative environmental~~
60 ~~impacts-accumulating-can accumulate~~ in water and soil and can become moderately toxic to organisms (Ellis et al., 2001;

4. l. 63 & 64: I think it has to be “Section” with “S” à please check!

We will leave this typesetting question to the editorial team and pay close attention during final proofreading.

5. l.70: “locations of industry” à maybe better “industrial areas” ?

Changed as suggested.

6. Section 2: l. 77: What does “approximately weekly” mean?

This means that this sampling is not strictly weekly but we try to collect samples at different times of day and on different weekdays. In addition, sampling is done manually and is coordinated with other maintenance tasks at the site.

7. l. 86: “ ...the sample loop is heated to approx. 200 °C...” - approximately?

This has been changed in the manuscript to not use the contraction (approx.) but approximately instead.

8. l. 99 “mud dauber” ...an insect screen, I guess?

We added that the mud dauber is also used as insect screen.

9. 1. 107ff: Are the sample inlets mounted in a separate heated box or are all parts in one single box? The description for the continuous instrument jumps a bit. Starting with the sampling procedure (flows, sample volume, desorption temperature) you continue with the unit set up (heated box, materials) and then return to the flushing procedure. The adsorption temperature is the same as in 2.2. (-80°C)? Streamlining this paragraph might improve the reading.

We improved this section regarding the reviewer's suggestion as follows:

110 2.3 Continuous in situ measurements

Continuous in situ measurements, with ambient air sampling every two hours, were started at TOB in May 2018. The air intake is a 3/8" stainless steel inlet line, located 12 m above the ground and, mounted outside a laboratory container with a goose-neck inlet. It uses a downstream diaphragm pump to continuously pull air from the inlet into the laboratory. ~~The 3/8" inlet, where the measurement system is located. This air intake~~ line is heated to 70 °C to avoid condensation and freezing. ~~This is done by heating cables installed at the inlet line and extra insulation surrounding the whole line.~~ To prevent the intrusion of particles, a mud dauber (Swagelok SS-MD-4), also used as insect screen, is installed at the open end of the inlet line. ~~In the laboratory, the inlet line is connected to one of five sample inlets, mounted at a heated box (80 °C), connected via 1/8" quick connectors (Swagelok). Inside the heated box, sample inlets are connected to a 10-position-selector (model EUTA-2SD10MWE, Vici Valco Inc., USA) with 1/8" stainless steel tubing. A Mg(ClO₄)₂ dryer (similar to the Goethe University laboratory setup; section 2.2) and a 4-port 2-position valve (model D4UWE, Vici Valco Inc., USA) are mounted inside the heated box and are connected via 1/8" stainless steel tubing. Directly before each measurement, the dryer and tubing of the system are purged and conditioned for 1 min at a flow of 100 mL min⁻¹ using the subsequent sample (air, calibration gas, etc.), bypassing the sample loop. Details are described by Obersteiner et al. (2016a) and are only briefly reviewed here.~~ Halogenated trace gases are analysed using a ~~gas chromatograph GC (Agilent 7890B), a TOF-MS (model EI-003, Tofwerk AG, Switzerland), and a preceding enrichment unit, which is similar to the enrichment unit used in the laboratory. Details are described by Obersteiner et al. (2016a) and are only briefly reviewed here.~~, using -80 °C for adsorption temperature, as well. For each measurement, approximately 500 mL of air are enriched in the sample loop at a sample flow of 80 mL/min. To determine the exact volume of enriched air, a mass flow controller (MFC; EL-FLOW F-201CM, Bronkhorst) and a pressure sensor (Baratron 626, 0-1000 mbar, accuracy including non-linearity 0.25 % of reading, MKS Instruments, Germany) are used. The sample loop is flash heated to about 220 °C for 120 seconds during sample desorption. Purified helium is used as carrier gas (quality 6.0, purification System: Vici Valco HP2). ~~Samples are dried using a trap to remove water vapour. Sample inlets are mounted inside a heated box (80 °C) and are connected via 1/8" quick connectors (Swagelok). The dryer, a 10-position-selector (model EUTA-2SD10MWE, Vici Valco Inc., USA), and a 4-port 2-position valve (model D4UWE, Vici Valco Inc., USA) are mounted inside the heated box and are connected via 1/8" stainless steel tubing. Directly before each measurement, the dryer and tubing of the system are purged and conditioned for 1 min at a flow of 100 mL/min using the subsequent sample (air, calibration gas, etc.), bypassing the sample loop.~~

10. 1. 133 : You neglect data for calibration intervals which deviate more than the weekly Is-precision, ok – I am just interested: Did you ever take in account/discuss to add the additional error to the uncertainty?

Yes, we have considered this. However, we decided against including such data points with poorer precision as this would e.g. be problematic when separating the data set into baseline data and outliers caused by local pollution. As this is not at the focus of this paper, we do not add a discussion on this.

11. Section 3: Figure 1: as I understand this figure is explanatory only, in order to describe the method how stable periods were defined. Therefore the plotted compounds are not mentioned. Nevertheless, the question bothered me while looking at the plot, what substances are plotted. Later on, in Figures 5 and 7 you show the similar plots again. I wonder, if you could combine those?

We added comments on that to clarify this. We agree, that Figures 5 and 7 show the same information for real data as is shown in Figure 1 for an artificial dataset. However, we think that for the reader it is more understandable to explain the procedure and the idea step by step and would therefore prefer to leave the figures as they are.

Figure 1. Schematic example [using a random data set](#) 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

215 [change in signal intensity leads to an incomparability with other compounds not by water influenced.](#)

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

12. 1. 181: Why did you choose the stability criterion to be 10%?

The value of 10% is somewhat arbitrary and was chosen as a compromise to have a sufficient number of measurement points while at the same time exclude data with a large measurement uncertainty.

We modified the manuscript as follows:

220 as grey data points in panel (b)). If more than one measurement has the same number of matching data points, the case with the lowest standard deviation is selected. In [our application, we have arbitrarily chosen a maximum deviation of 10% as a selection criterion, as it allowed to retain a sufficient number of measurements while still eliminating data which would have particularly large uncertainty. Depending on the stability of the instrument and the desired results, different criteria could be chosen. Allowing for larger deviations would result in retaining more data with larger uncertainties, while applying a more](#)
225 [stringent criterion would result in a dataset with less data yet likely also lower uncertainties.](#) In panels (c) and (d) the evaluation

13. l. 195 ff. I had to read this sentence several time. Just to be sure: MAPE is the difference “ $A_{\text{measured}} - A_{\text{odr}}$ ”, with “ A_{odr} ” being the ODR fit of peak areas forced through the origin

We clarified this as follows by adding also the formula used to calculate MAPE:

235 the mean absolute percentage error (MAPE) ~~MAPE is calculated using the~~, defined as:

$$\frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{O(t) - F(t)}{O(t)} \right| \cdot 100 \quad (6)$$

where n is the number of observations. $F(t)$ is the predicted data as orthogonal distance regression fit forced through the origin and ~~the deviation~~ $O(t)$ is the data of the observed peak areas ~~to those determined from the fit~~. The r^2 and the MAPE

14. l. 210: “Using HFC-227ea and....” You already present a single result of the analysis that follows in the next paragraph. I found this is confusing. From my point of view you can delete this sentence. Figure 5 (&7): I understood that you derived stable periods from calibration measurements as well. However, in the caption of Figure 5 you mention “data of the weekly flask sampling measurements“. Please clarify.

We changed l. 210 as follows, to make the section more clear. Using this sentence, we want to point out the importance of the selection of appropriate substances to receive highest possible amount of data.

the number of measurements after selecting the periods of constant rRF should remain as high as possible. ~~Using~~ E.g. using HFC-227ea and HFC-245fa as evaluation substances in combination with HFC-143a as main reference substance or using HFC-245fa as a main reference, more than half of the calibration measurements are excluded, as shown in Fig. 5, panels (d) and (h). As this leads to a significant decrease in the number of air measurements for which an indirect calibration value can
265 be derived, these substances are also less suitable as reference substances.

“Data of weekly flask sampling measurements” means that rRF is calculated using the calibration measurements, used during the flask sampling measurements. We clarified this as follows:

Figure 5. Illustration of ~~data selection of data of the~~ constant rRF data selection, using the standard calibration measurements during the weekly flask sampling measurements, using two different main reference ~~substance~~ substances (HFC-143a (a-d), HFC-245fa (e-h)) and an

15. l. 221ff: “To quantify the differences...” I do not understand this sentence, please explain. As a consequence please also explain panels d and h in Figures 5 and 7. Not clear to me.

We clarified this as follows:

and test substance. To quantify the ~~differences between the selection of data of main reference and test substance via main reference substance and precision loss between direct calibration and calibration via~~ an evaluation substance, we compared
275 the relative standard deviations of the resulting ~~filtered data sets~~-dataset as follows: (i) the rRF_{test} -dataset, applying the 10% filter criterion directly (Fig. 5 (c) and (g)), and (ii) the rRF_{test} -dataset, using the data points which are selected via the residual rRF_{eval} -data points applying the 10% filter criterion (Fig. 5 (b) and (f)). This is shown for all substance combinations in

16. l. 229ff: “In summary...” These are the characteristics you expect from your relative Response factor to achieve a good calibration and which set the frame for your checks you presented in the preceding lines. Might be good to have it earlier in the text?

In the preceding paragraphs of the subsection we have discussed several aspects of the variability of the rRF. We therefore think that it makes sense at this point to briefly summarize the characteristics that are needed right before discussing our choice of substances. At this point we summarize the characteristics discussed in the previous paragraphs of this subsection.

17. l. 237: “Using HFC-125 as evaluation substance with HFC-143a, the difference standard deviations of the mean rRF selected via the test substances and selected via itself ranges between 1 and 10 %.” Do you mean “different”?

Yes, we meant different and this has been altered in the revised manuscript.

18. l. 243: “In this test case, mole fractions of HFC-227ea shows the best correlation ...”
“show” instead of “shows”.

Altered to “show”.

19. l. 249: You end this paragraph with a statement about the importance of having a constant sensitivity. You take up this point but then do not really discuss it. It would be nice to have all the presented results of 3.2.2 “wrapped up” at this point in final short summery

We think that this has been discussed in the previous paragraphs and refer to the summarizing sentence that is an introduction into the final parts of subsection 3.2.2 (cf. comment 16. above). We therefore prefer to end this subsection with the statement about the importance of a constant response factor.

20. Table 2: With standard deviation you mean the standard deviation of the average relative difference? Is it this really necessary to have this table or could you add this information into Fig.6. ?

Standard deviation here indeed refers to the standard deviation of the relative deviations of the data points. An explanation has been added to the table caption.

Adding more information into Figure 6 would result in an even more crowded figure. We think that therefore a table is better suited and, in addition, it facilitates the comparison of the numbers for the three compounds.

When reviewing the data, we discovered an error in the numbers in Table 2 and this was corrected.

and HFC-245fa poorer results with $r^2 = 0.79$ and $r^2 = 0.63$, respectively, are obtained. The ~~relative average differences of the direct and~~ MAPE, where $F(t)$ is now defined as the indirect calculated mole fractions, and $O(t)$ defined as direct calculated mole fractions (cf. eq. 6) are given in Table 2. ~~Table 2 shows also the standard deviation of the relative deviations between indirect and calculated mole fractions, to show the spread of the differences between direct and indirect calculated mole fractions.~~ The rRF of HFC-245fa as evaluation substance and HFC-143a as main reference substance has also less than 50 % of selected data within the 10 %-filter (cf. Fig. 5), which means that the calculation is applied to a large portion of data for which the criterion of a constant rRF was not met. This underlines how crucial the assumption of constant instrumental sensitivity is for the indirect calibration method.

Table 2. ~~Average relative differences~~The mean percentage error (MAPE) of directly and indirectly calculated mole fractions ~~and standard deviations of the relative deviations of each data point (direct vs. indirect)~~ for the whole air flask sample GC-MS measurements for the period from October 2013 to December 2018 (cf. Fig. 6). As main reference HFC-143a, as evaluation reference HFC-125 is used.

| compound | av. rel. difference MAPE | standard deviation |
|-----------|-------------------------------------|----------------------|
| HFC-32 | 21.37,5 % | 20.96,4 % |
| HFC-227ea | 19.311,1 % | 13.912 % |
| HFC-245fa | 13.28,8 % | 99,0 % |

21. *Figures 8 and 6: When you calculated the amount fractions via the indirect way you applied the method for data in the stable periods (as derived in Figure 5 and 7 respectively). The constant rRFs you used to “attach” the test substances to the reference substance are determined using the average rRF in the calibration measurements also during the stable times? This is not clear to me? Why did you use the measurement precision as error bars? How does this reflect the uncertainty factors of the different calibration methods?*

It is correct that average rRFs are only determined based on data from periods of stable rRF, as described in section 3.2 Method evaluation.

It is correct that in Figure 8 the error bars represent the standard deviation of the measurements of one day. The data points shown represent daily averages. Therefore, these error bars should not be taken as a measure of the absolute error but rather as an atmospheric variability. We specified this in the modified figure caption:

Figure 8. As Fig. 6, but for the continuous in situ measurements at Taunus Observatory using GC-MS and for the following compounds: HFC-125, HFC-143a, and HFC-245fa, calculated directly (yellow symbols) and indirectly (~~light~~-blue symbols). Here, HFC-152a was used as main reference substance, PFC-318 as evaluation substance to select data with constant rRF. For reasons of simplified illustration daily means are shown. The error bars indicate the standard deviations of the measurements of one day and thus reflect the daily atmospheric variability rather and do not include systematic errors due to the indirect calibration method.

22. *Figure 6: What happens if you omit the two single data points in January 2018 for HFC-32? Any idea what happened here?*

Omitting the two high values in the case of HFC-32 in Figure 6 results in a value of r^2 closer to 1 and a slope closer to 1. Because we do not have any indications of a technical issue causing these outliers, we did not remove them. The correlations are used as a pre-selection criterion. As a large number of outliers worsens the correlation, the example of HFC-32 shows the necessity for such a criterion.

23. *Figure 8: What happens to the regression lines when you omit the high amount fractions?*

For data shown in Figure 8, omitting the high values leads to smaller values of r^2 . In fact, this shows the limitations of the method as discussed in the conclusion section. The method is not well suited for the detection of small variabilities or of small trends due to its rather large uncertainty.

24. *Tables 2+3: Do you use this deviation to derive the uncertainty of the method?*

Yes. We use this deviation to quantify the uncertainty of this method. Because we cannot find other ways to compare the unknown data points with known data points, we use this approach (using the test datasets) to find an appropriate quantification for the determination of the uncertainty.

25. l. 262: “This is caused by long-term drifts...” What does this tell you about the possible errors arising from differences between the evaluation and test substances, even though you have filtered out periods with a larger variability?

This confirms that the method is not applicable for small long-term trends, as we point out in the conclusions section.

26. Section 4: Figures 9 and 10: maybe you keep the zoomed in plot, only? What do the error bars represent?

To visualize our whole data records, we prefer to leave both plots in the paper. We explained the error bars as follows:

Figure 9. Time series of mole fractions of HFO-1234yf (a), HFO-1234ze(E) (b), and HCFO-1233zd(E) (c) at TOB. Symbol shape indicates flask sampling (diamonds) and in situ measurement (circles). Grey symbols represent mole fractions below detection limit (cf. Table 2 and 4). Data of weekly flask sampling are indirectly calibrated mole fractions before January 2018 and directly calibrated values after. For the in situ measurements, indirectly and directly calibrated mole fractions are indicated by color (orange and yellow). Note the change in x-axis scaling after January 2018 because of increased data frequency. [Data of in situ measurements show daily means. Errorbars for weekly flask sampling show the 1 \$\sigma\$ standard deviation of the measurements for one sampling. Errorbars for the in situ measurements show the standard deviation of the average of measurements over one day.](#)

27. l. 291: “These larger amounts could be....” So, this is the effect of non-linearity or a larger integration error for the small calibration peak?

Up to now we have not found any indications of non-linearity for the investigated H(C)FOs but it seems likely that the large uncertainty results from the size of the peak. We have reworded this for more clarity:

large deviation with maximal 44 % for the data in 2018. These larger [amounts-deviations](#) could be caused by the [low-amount](#)
360 [small peak](#) of HFO-1234ze(E) in the calibration gas used from May 2018 to March 2019. The mean values of both substances,

28. Conclusion: You have presented methods to derive the best possible reference substances for your indirect calibration and you evaluate this indirect calibration procedure regarding its performance. It would have been nice to see an assessment of how uncertainties arising from the determination of the rRF and evaluation of stable periods (by the evaluation) are reflected in the results of the test compound analysis. E.g. do expected errors match with observed differences between indirect and direct calibration?

In this work we focused on the development of this approach and on demonstrating its application to atmospheric measurements. To disentangle the various sources of uncertainties, further data evaluation would be necessary, for example applying the method to a larger number of substances. In addition, we still need a better understanding of the causes of the periods of unstable rRF.