

Reply to RC 2

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. *General comment: The manuscript contains a large number of figures and perhaps one or two of those could be relegated to SI ?*

We have gone through the manuscript and considered reducing the number of figures. However, we found that all figures are necessary to follow our evaluation approaches and therefore we prefer not to reduce the number of figures.

2. *Abstract – line 7: “thus” can be omitted in the sentence.*

Done.

3. *p. 2 line 36: Typo in spectrometry (spectrometry).*

Done.

4. p.2, line 53: *It is incorrect to say “TFA is known to cause negative environmental impacts”. Large concentrations of TFA will do that, but the authors should consult the references they cite themselves, especially Solomon et al., for a precise characteristic on this matter, what impact the current and predicted future levels of TFA in the environment will actually have.*

Changed this as follows:

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;
Russell et al., 2012; Solomon et al., 2016). ~~In recent studies it seems the TFA amount formed from the mentioned substances~~
~~in the troposphere may be too low to cause negative effects on human health (Solomon et al., 2016). But there is a necessity,~~
~~to investigate these sources of TFA in more detail, as it is done in Solomon et al. (2016) or Freeling et al. (2020). Our data and~~
65 ~~approach can be a helpful additional tool for those investigations, and the exploration of seasonality, or temporal and spatial~~
~~trends.~~

5. P4. line 115. *Mole should be capitalized ion the beginning of the sentence.*

Done.

6. p.6 line 159-161: *This sentence should be rewritten for clarity. It would benefit form some commas and perhaps start out with “Using equating 3, the rRF for the species of interest, which is not.....*

Changed as suggested:

calibration method. ~~The Using Eq. (3), the rRF for the species of interest which is not present in the standard, relative to a~~
190 ~~compound which is detectable in the standard, can then be derived from measurements of another sample which has detectable~~
~~amounts and known mole fractions of both species using Eq. (3).~~

7. p.6, line 165: *replace “should” with “is assumed to”*

Done.

8. p.7, line 173: *I guess the selected compounds listed in table 1 could be termed “a training” set for the method. As such, the authors, and other researchers, will adopt similar or dissimilar training sets, for the technique to be “calibrated’ on for application to their datasets.*

We thank the reviewer for this useful suggestion. We added the following:

Table 1. System precision (1σ) of the investigated substances treated as a training set of the TOF-MS used for the weekly whole air sampling (prc (TOF_Lab)) and of the TOF-MS used for the in situ measurements (prc (TOF_in situ)) and their calibration scales.

of the rRF are and to determine periods of low variability of the rRF , we have investigated the temporal change of rRF
205 for the combination of selected compounds listed in Table 1, which we call a training set here. Substances in Table 1 were

9. p.7, line 182: 10% - this this value arbitrarily chosen? Why not based on a statistical parameter such as sigma(s)?

Yes, it is arbitrary chosen. We decided to use this 10 % criterion instead of sigma(s). But it could be adjusted in further investigations. We added a comment on that:

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 the retention 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 stringent criterion would result in a dataset with less data yet likely also lower uncertainties. In panels (c) and (d) the evaluation

225

10. Figure 4 caption: replace “used cylinders” with “used calibration gas”. All figures could in general benefit from being made color “agnostic”. Several of them, e.g. fig 6 and 8, are only legible in color print out, whereas there are no limitations on symbols shape that dictates the necessity of using colors.

Changed wording as suggested and we changed the colors and symbols of Figures 6 and 8 to follow the suggestion of the reviewer. They are now more “agnostic”, yet are still best viewed with color options.

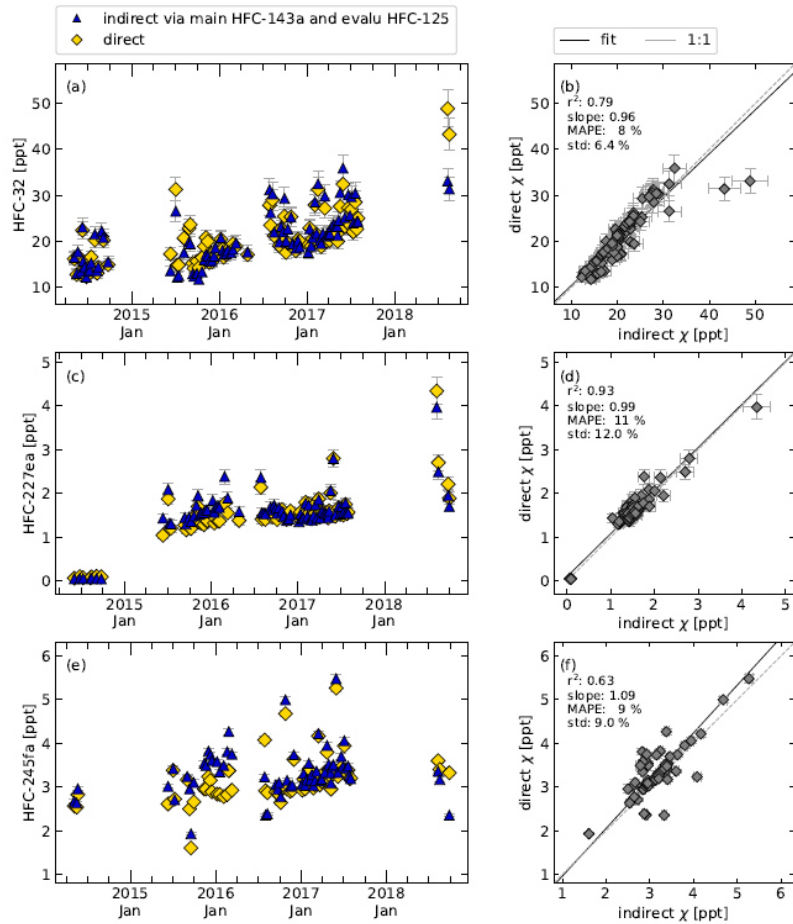


Figure 6. Time series (left) and correlations (right) of the mole fractions of HFC-32, HCFC-227ea, and HFC-245fa calculated directly (yellow symbols) and indirectly (light-blue symbols) for the weekly flask sample measurements. HFC-143a was used as main reference substance, HFC-125 as evaluation substance to select data with constant rRF . Error bars, which indicate the measurement precisions, are included but are often smaller than symbol size.

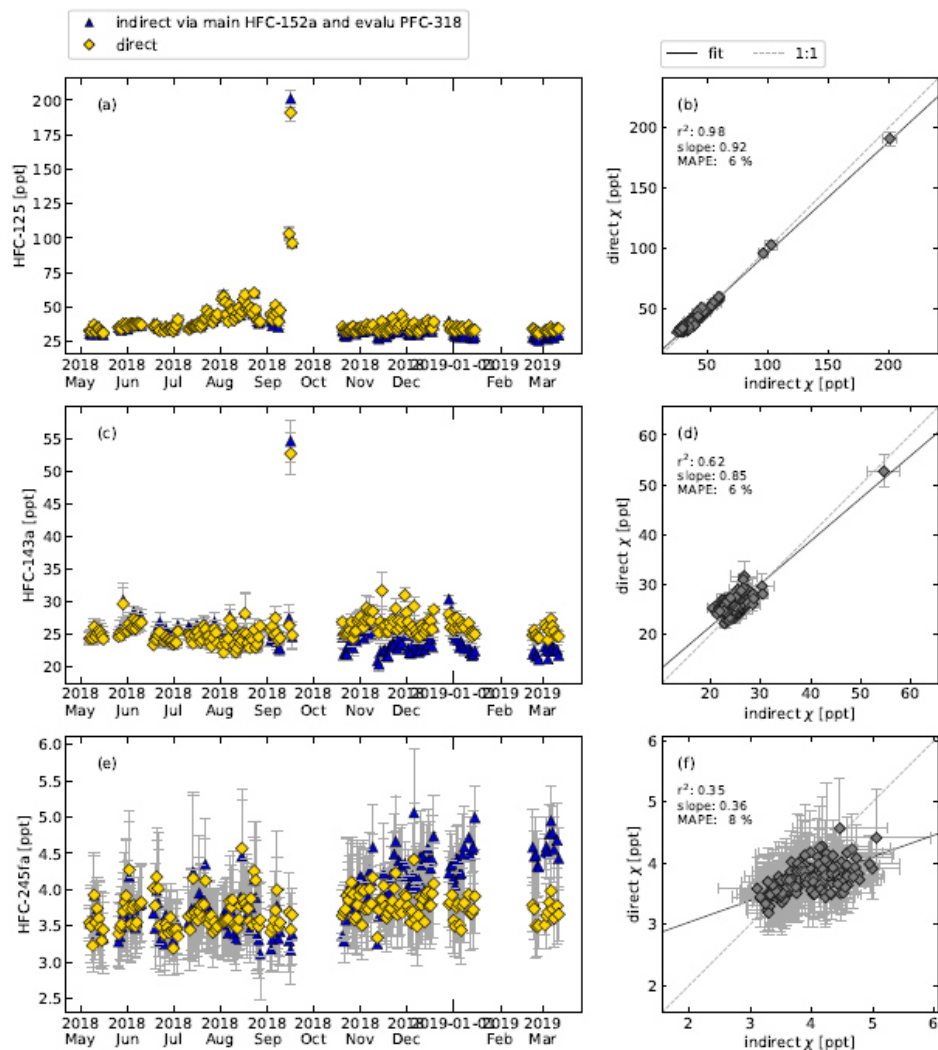


Figure 8. As Fig. 6, but for the continuous in situ measurements at Taunus Observatory using GC-MS and for the following compounds: HFC-125, HFC-143a, and HFC-245fa, calculated directly (yellow symbols) and indirectly (light-blue symbols). Here, HFC-152a was used as main reference substance, PFC-318 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.

11. P.15, line 254-255: Please comment on the fact that a larger fraction of data from the in-situ measurements than from the flask measurements were selected. Was this expected? Any predictable reason behind this outcome?

We added a comment on that.

from the in situ measurements than from during the weekly flask sample measurements. This could be due to the shorter time period covered by the in situ data and also due to the continuous measurements in contrast to flask measurements in the laboratory, where the instrument is in standby for longer time periods. HFC-152a as main reference has a high overlap within

12. P.16, line 262: What does “partly different” mean here? Could a different description be used?

We clarified that:

Especially shorter-term variations are well captured, while long-term trends between the directly and indirectly calculated mole fractions ~~are partly different~~ sometimes show systematic differences between the directly ~~determined and the indirectly determined data~~ and indirectly determined methods. This is caused by long-term drifts in the rRF and shows clearly that the indirect calibration measurement should only be applied when investigating very large long-term trends when no directly calibrated measurements are available. The average relative differences are given in Table 3.

13. p. 16 line 275, Suggest replacing detectability with detection frequency.

Changed that as suggested.

14. p.16, line 278 and 281: inset “ weekly” before “mole fractions”.

Added that.

15. p.18, line 285-287: How would this indirect retrospective method likely work out trained on a data set like that collected at Jungfraujoch – i.e., a setting with largely clean background air? Any ideas if it would work there as well?

We think that the method would work out probably similar. We have to mention, that the long term detection of minor and small changes is not the strength and not the aim of this method. We decide not to change the manuscript to point that out. We added a comment on that few lines before (cf. answer on point 12).

16. p. 19, line 295-297: Does this mean that indirect values obtained from datasets involving flask measurements and in-situ measurements, respectively, shouldn't really be directly compared? I.e., if the variability is quite unpredictable, which sample dataset is "generally" more likely to produce good indirect values.

No, this does not mean that such data sets should not be combined. It is just again due to the possibility of long-term systematic trends in rRF. We added a note on this:

365 unequal time distribution of the data. ~~This could cause also~~ Also, this could cause the other deviations between the in situ and whole air flask measurements. Such deviations are to be expected due to possible long-term drifts in rRF, again emphasising the point that the indirect calibration method is better suited to investigate short-term variability in ambient air measurements than for the detection of long-term trends. However, typical mole fractions are also in between typical mole fractions observed

17. p.21, line 311: "sufficiently" is probably a better word here than "rather".

Changed as suggested.

18. p.21, line 311-312: regarding "reference species with similar retention times" – has the sensitivity to the retention-times been tested? Is it just assumed "likely"- I'm not saying that this is not a good assumption, just wondering if the authors have tested this – if not, this should be an easy test within the GC-TOFMS data sets.

We added a comment on that:

using species ~~whose concentrations with concentrations that~~ are known in the calibration gas. We suggest that it is useful to use an evaluation substance to select periods when relative responses of the measurement system are rather stable. Further, it is likely that using reference species with similar retention times as the target species provides more stable results-, which should be investigated with a larger number of training substances. The training dataset used in this work could not confirm
385 that. By analysing correlations and variabilities of the relative responses, we identify the combination of a main reference and

19. p.21, line 317: "... retainage a sufficient number of measurements". Sufficient here means a very low number? Ref. Table 4 where e.g. 2018 has 3-6 observations?

We clarified that as following:

and the evaluation substance are stable within 10% in the analysis. A good combination of reference and evaluation substance should thus yield small deviations between direct and indirect calibration for a wide range of compounds, while also retaining
390 a ~~suffieient-number~~ maximal fraction of measurements based on the filter criterion of maximum deviation in relative response factor, if possible.

20. p.22, line 325: *These quoted different are much lower than what shows up for the annualized values in the tables. Is the large discrepancies for the annualized values not an issues since those are what are often cited?*

In the summary, we give the averaged deviation by comparing the data points calculated directly and indirectly of each individual measurement, as shown in tables 2 and 3. Larger annual discrepancies only occur in the comparison of HFO-1234ze(E) in table 5, e.g. for 2018 annual mean direct: 1.11 ppt, annual mean indirect: 0.63 ppt. Whereas HFO-1234yf shows annual deviations between 5 and 13 % (e.g. for 2018 direct: 0.63 ppt, indirect: 0.66 ppt).

We added the text:

~~time-of-flight mass spectrometric-GC system with TOF-MS~~ detection. This ~~data-set-dataset~~ has been evaluated for the time period from May 2018 to March 2019. Comparing the data points of each measurement, calculated directly and indirectly, we found the following averaged relative differences of mole fractions for the investigated substances: For the long-term flask data, we find relative differences between directly and indirectly calibrated mole fractions of different gases ranging between 400 ~~13 and 218~~ and 11 %. For the in situ data, differences between directly calibrated and indirectly calibrated mole fractions ranged between about ~~15 and 256~~ and 23 %.