

Atmos. Meas. Tech. Discuss., referee comment RC3
<https://doi.org/10.5194/amt-2020-488-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2020-488

Anonymous Referee #3

Referee comment on "An indirect-calibration method for non-target quantification of trace gases applied to a time series of fourth-generation synthetic halocarbons at the Taunus Observatory (Germany)" by Fides Lefrancois et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2020-488-RC3>, 2021

Review for manuscript by Lefrancois et al., 2021.

The manuscript by Lefrancois et al. presents a method to derive quantitative information from the measurement of trace gases made by preconcentration, GC-MS, even in cases where the substance in question is absent from the calibration gas. This method answers a growing need for making both retrospective screening and non-target screening quantitative. The method is then applied to retrospective screening of emerging HFOs, and therefore reconstructs quantitative molar fractions values for the period 2014-2018, for the site of Taunus Observatory in Central Germany. We can expect that the presented method will be widely applied in the future, to accompany the deployment of high mass coverage, high mass resolution instruments at monitoring stations.

The paper is generally well organised and Figures well chosen. It is evident from the data presented that the method has been well tested, and the results are of good scientific value. However, a few minor improvements would help future readers. Reformulation of a few sentences is needed (specific suggestions hereafter). Also, where possible, it would be very interesting to mention potential causes for the rejection of about half of the data points.

Comments are detailed hereafter.

I would suggest to use a title that contains the main purpose of the article, which is to make non-target screening quantitative. Usually, articles with "non-target screening" in their title contain discovery of new substances, which is not the scope here. Suggestions for title: "Quantitative non-target"

l. 8: I would simplify the grammar: "This archive can be used if", or even "This archive can be used for retrospective screening"

l. 11: "or the amounts in the calibration gas may not have been quantified."

Introduction, l. 18-20: "The application of the indirect calibration method on several test cases can result into accuracies around 13% to 20 %. For H(C)FOs accuracies up to 25% are 20 achieved." I would be good to reformulate these sentences to really convey the meaning that low values represent a better accuracy. Maybe you can replace the word "accuracy" by "uncertainty" here: "The application of the indirect calibration method on several test cases can result into uncertainties around 13% to 20 %. For H(C)FOs, of particularly low mole fraction values, uncertainties up to 25% are observed."

l. 26 "which are part of or affiliated to"

l.31, maybe: "is not well covered"

l. 33: "by the installation"

l. 35: maybe you want to specify the mass range coverage of the TOF instrument (minimum and maximum measured masses).

l. 45-47: may you can consider if you would like to leave out "HFC-1234yf" and "HFC-1234ze(E)". The HFC nomenclature is actually not made for compounds with a double bond.

l.51 : "have no ODP": "have an ODP value of zero." "have no ODP" suggests that the computation of the ODP is impossible.

l. 57: the magnitude of what? Amplitude of annual cycles, magnitude of mole fraction of pollution events?

l. 80: "and each pair of measurements is bracketed"

l. 81: "range of parts per trillion (ppt)": range of picomole per mole, pmol/mol or hereafter part per trillion (ppt)"

l. 102: add comma: "For each measurement, approximately"

l. 117: you can leave out the sentence about calibration scales, it is already mentioned l. 92-94.

l. 139 : "before calibration standards containing measurable amounts of these substances were used".

l. 140: tense concordance, not sure, check with native speaker. "When these compounds were detectable in ambient air, the peak areas could not be converted to mole fractions using Eq. 2 because neither numeric values for A_{cal} nor rR were available."

l. 141: you surely mean: "between another compound which is measurable in the standard"

l. 144: "that means that the ratio of signal per amount of analyte for the two compounds is constant with time." I'm not sure about the meaning of this sentence. We know that the response of a MS instrument may vary strongly over time, for example the instrument response increases after source cleaning. However what is important here is that the instrument response behaviour should vary similarly over time for all substances, as you clearly write afterwards. I would rephrase as: "Ideally, the sensitivity of the analytical system for two different species should behave similarly over time. In such a case, the ratio of responses R of two given species should be close to constant."

l. 146: "this ratio should be the same for any sample.": maybe too general. Suggestion to write more specifically: "this ratio should be constant over time for any chosen pair of compounds".

l. 155: "It must be stable over time". Check entire manuscript.

l. 164-166: meaning not clear. A non-stable sensitivity does not necessarily imply a non-stable relative sensitivity, this is something you are going to investigate next. Suggestion to rephrase: "The methodology outlined in 3.1 is based on the assumption of a constant rRF in Eq. 4. In reality, the absolute sensitivity of a mass spectrometer is known to vary over time, in particular after tuning the mass spectrometer or after modifications of the

analytical system such as replacement of filaments, columns or sample loops. It is therefore an open question whether changes in the relative sensitivity rRF should also be expected or not. Thus, to evaluate [...]"

l. 169: "need to separated": "need to be evaluated separately".

l. 181-185: difficult to understand, suggestion to rephrase: "To identify periods of stable rRFevalu for a given pair of compounds, timeseries of rRFevalu are reviewed. To do so, for each measurement or data point of rRFevalu in the timeseries, we compute the sum of other rRFevalu data points that do not deviate from the chosen data point by more than 10%. The data point with the highest number of matching data points is used as a reference (shown with red circle in Figure 1, panel (b)) and all data points that fall outside the 10% interval are excluded (shown as grey data points in panel (b)).

Note: I would not use "independant measurement", since the measuring instrument is the same of course the results are not fully statistically independent, and we actually need the results not to be independent for this method to work.

To make it more clear, on Fig. 1 please mark with e.g. a red circle the data point that was selected as most likely rRF value.

Table 1: add bibliographic reference to all scales where needed.

METAS-2017: Guillevic et al., 2018 (ok, already done).

EMPA-2013: for HCFC-133a: Vollmer, M. K., Rigby, M., Laube, J. C., Henne, S., Rhee, T. S., Gooch, L. J., Wenger, A., Young, D., Steele, L. P., Langenfelds, R. L., et al. (2015), Abrupt reversal in emissions and atmospheric abundance of HCFC-133a (CF₃CH₂Cl), *Geophys. Res. Lett.*, 42, 8702– 8710, doi:10.1002/2015GL065846.

EMPA-2013 for HFOs: Vollmer et al., *Environ. Sci. Technol.* 2015, 49, 5, 2703–2708.

SIO-05, SIO-14: Prinn et al., *J. Geophys. Res.*, 105, 17,751-17,792, 2000, and Prinn et al, *Earth Syst. Sci. Data*, 10, 985–1018, 2018.

l. 195: you probably need "the" in front of all "MAPE", check through the manuscript. I would add the equation for the computation or a reference (e.g. the Wiki page).

l. 199: "Except for HFC-227ea"

Section 3.2.1, general question: could you find explanations for the outlier rRFevalu data points?

l. 200-201, I would try to reformulate in an easier way. E.g.: "To test which pairs of substances produce the highest correlations, all possible pairs of substances have been tested. The obtained values for r^2 and MAPE are shown in Fig 3".

l. 208, typo: "all cases where HFC-152a is involved." Also, it seems to me more to be a drift in the rRF value, that started before the change in standard tank, and stabilised after some runs of the new standard. Such a drift (albeit much smaller) can also be seen in the HFC-125 data points. So I'm really not sure that you can link this for the standard tank change. I would remove this sentence.

l. 211: maybe you can comment on why HFC-227ea and HFC-245fa? HFC-227ea seems logical to be a bad one, as its measurement standard deviation given in Table 1 is one the highest. However why HFC-245fa? Or, alternatively, you can explain later why some are good ones?

l. 220-222: "To quantify the differences between the selection of data of main reference and test substance via main reference substance and an evaluation substance we compared the relative standard deviations of the resulting filtered data sets." I don't understand this sentence. Please clarify. You may also want to cut into smaller sentences. Maybe, adding the equation you use will help to understand what you compute here. Usually there are two quantitative values to characterise a result: its standard deviation, which reflect the random noise, and the average difference between two values (usually a test value and a reference value), which is a systematic bias. A bias not equal to zero means that the method causes a systematic error. Now based on Fig 5, maybe what you want to express here is a precision loss, that you express via the difference in standard deviation? If this is really the case, here is my suggestion:
"To quantify the precision loss between direct calibration and calibration via a transfer substance, we compare the relative standard deviations of the resulting filtered data sets.", or something similar.

Another important quantity to evaluate is if your method creates a bias or not? i.e. what is the average value of the distance (or difference) between the true and reconstructed value? It should be (close to) zero to show no bias. (cf see below comment on Table 3)

l. 237: if you mean precision loss, use: "the difference between the standard deviations".

l. 241: "As test cases to apply the indirect calibration method, we chose..." or "As test cases to be applied the indirect calibration method, ...".

l. 243: "mole fractions of HFC-227ea show..."

Table 3: average relative difference: this is your metric for the bias, right? Please write the equation somewhere in the text (e.g. around l. 245). Also: usually, if the bias or systematic offset value is within the 2 sigma standard deviation, it means within uncertainty, there is no bias. This is an important point to show here. But in Table 3, the "av. rel. difference" value is systematically more than the value of "standard deviation". Can you comment on this?

l. 275: typo: "HFO-1234yf"

l. 276: concordance of tenses, "increased continuously up to 100%"

l. 311-312: "Further, it is likely that using reference species with similar retention times as the target species provides more stable results." Can you give an example here? No retention time data are provided.

l. 313: "good results" is subjective. Maybe use a quantitative value instead, e.g. "which yield the minimum number of rejected data points".

l. 330, typo: "is the measurement", "which are expected".

Your data show a rRF that is mostly not stable over time. Can you discuss the possibility to use a running-mean rRF value over time, instead of assuming a constant value over a short time period? Also, at least for some time periods, could you assign a (hardware?) cause to the non-stable rRF?

Figure 5, legend: "Illustration of data selection for the weekly flask sampling measurements..."