

Interactive comment on “ACTRIS non-methane hydrocarbon intercomparison experiment in Europe to support WMO-GAW and EMEP observation networks” by C. C. Hoerger et al.

C. C. Hoerger et al.

corinne.hoerger@empa.ch

Received and published: 20 January 2015

J. Rudolph (Referee) rudolphj@yorku.ca Received and published: 25 November 2014
The paper presents results from an intercomparison exercise for European laboratories conducting atmospheric measurements of volatile organic compounds (VOC). Most of participating laboratories are part of the Global Atmospheric Watch (GAW) or the Aerosols, Clouds, and Trace gases Research InfraStructure (ACTRIS) network and conduct VOC measurements in the atmosphere either as monitoring activity or as part of major field campaigns. VOC, as pointed out by the authors in the introduction, play a major role in the chemistry of the atmosphere and consequently there is substantial

C4647

interest in VOC monitoring data for scientific and regulatory purposes. Understanding the quality of the measurements is essential for any meaningful use of VOC data which have been and will be produced by these laboratories. Consequently the results of this intercomparison are of interest to any potential user of VOC data produced by these laboratories. The intercomparison consisted of sending around two samples, one artificial mixture and one air sample to the different laboratories. The participating laboratories analysed these samples for VOC molar fractions and reported these results together with some performance indicators for the quality of measurements, such as repeatability, blank values and uncertainty estimates. The paper presents the results of this exercise and also discusses many details of the various sources of uncertainty and bias for measurements which significantly differed from the target values. Although most of the 20 participating laboratories use gas chromatography for VOC analysis, the instruments used differ in many details and represent a substantial and important subset of state of the art VOC measurement techniques. Therefore the paper potentially is also relevant for readers who want to get insight into the problems and difficulties that impact the quality of VOC measurements. I also appreciate that the authors provide details of the measurement methods used by the different laboratories and present the results in a way that allows identification of the laboratories with data provided by the individual laboratories. Such openness clearly enhances the value of the paper for all users of NMHC data provided by any of the participating laboratories. However, in its current form the paper is not without problems and of limited value for readers not involved in ACTRIS or GAW. i) Although the procedure used is that typical for an intercomparison, the evaluation of the results is simply the comparison of measurement results with the composition of two reference samples. This evaluation is mainly based on the ACTRIS and GAW quality objectives as described in 2.5. The composition of the reference sample was determined by three laboratories and the molar fractions and their errors were assigned using a simple statistical evaluation (2.3). Most of the discussion of uncertainties seems to imply that the uncertainty of the reference samples has been accurately determined by this procedure, the true

C4648

accuracy of the reference sample is not evaluated at all.

The “true accuracy” is hardly assessable as we have to rely on some assumptions: 1. The NPL standard defines the scale and uncertainties of NPL are correctly stated. 2. Gaussian error propagation is valid for the individual uncertainty determination of each reference laboratory. 3. The combination of 3 independently determined mole fractions by assigning laboratories with an appropriate combination of internal and external errors reliably describes the uncertainty relative to the scale. Corresponding clarifications have been added to section “Determination of assigned values. . .”

There are a number of reasons why uncertainties derived from a statistical evaluation may not represent the true accuracy. First of all, a comparison between three laboratories is hardly sufficient to determine the accuracy of the VOC molar fractions in the reference sample without a detailed evaluation of the independence of the analytical methods used. A simple example is potential peak overlap. Are the columns used for separation sufficiently different in polarity to assure that the results from the different laboratories do not suffer from similar interference problems? The same question applies to other components of the analytical systems used for determination of the reference sample composition. HPB and Empa both used Al₂O₃/KCl, however, with different temperature programs and flow rates, and the WCC-VOC used Varian CP-Silica PLOT parallel with Varian VF 1 MS, thus sufficiently different separation methods were used. With the two columns of WCC-VOC peak overlaps can be circumvented. Where a peak overlap occurred the uncertainty was increased accordingly. The preconcentration systems are substantially different, as outlined in Table 4a-4b, with e.g. custom built adsorptive sampling at -45°C (Empa), LN₂ cryo-sampling on glass beads (DWD), and adsorptive sampling at -120°C (WCC-VOC). In result, no discrepancies beyond the estimated uncertainties were determined. Furthermore, the NPL mother mixture (100 nmol/mol) was filled from the same tanks as the NPL certified 2 nmol/mol standard used by DWD. The dilution of the 100 nmol/mol mixture to roughly 1 nmol/mol (NMHC_N2) yielded exactly the same relative composition as the certified 2 nmol/mol

C4649

mixture indicating consistency to NPL scale within the analytical uncertainty as stated in Table 5.

Furthermore, the three laboratories used the same standard (NPL) to determine the composition of the reference samples. Consequently uncertainty in the molar fractions assigned to the standard will not be reflected in the statistical comparison between the results of the three laboratories or any other laboratory that used a NPL standard or a standard that is traceable to NPL. Indeed, it is interesting that in subchapter 3.3 it is stated that “The systems using a NPL standard for direct calibrations (Table 4) generally exhibited a good performance”. The NPL defines the scale in GAW NMHC measurements – it is the Central Calibration Laboratory. NPL compares the NMHC scale and associated uncertainty in key comparisons with other NMI (National Metrology Institutes) (Grenfell, R. J. P., M. J. T. Milton, A. M. Harling, G. M. Vargha, C. Brookes, P. G. Quincey, and P. T. Woods (2010), Standard mixtures of ambient volatile organic compounds in synthetic and whole air with stable reference values, *J. Geophys. Res.*, 115, D14302, doi:10.1029/2009JD012933 and references therein.). Furthermore, the FID-carbon response factors indicate excellent consistency between the various NMHC. Thus, the reference to NPL scale with the stated accuracies is assumed as justified basis. This can be expected since uncertainty in the calibration standard will cancel when comparing measurements using the same calibration standard.

The authors intended to make the difference between systems directly calibrated by NPL and those relating their calibration to a whole air working standard that in turn had been related to a NPL standard. It was not surprising that the 2-step calibration was not as good as a direct calibration.

For measurements based on a calibration standard that is independent of the NPL standard, uncertainties in both calibrations will propagate in to the difference between measurement and reference sample and it is therefore expected that measurements using an NPL independent calibration will on average show a somewhat larger uncertainty.

C4650

The reviewer is right in this point; however, it only concerns measurements by 4 laboratories (FZJ, PUY, MHD, SMRII) which are related to different standards (Table 4). The corresponding results reported by these laboratories differ substantially from the reference, with deviations typically much larger than the combined stated uncertainty of the calibration standard. Furthermore, these results are, except for PUY, related to Apel-Riemer Inc. standards. If the systematic offset between different scales exists, it should result in systematic deviations from the reference values. But this was not observed here (Fig. 2, revised version). Thus, the measurement quality of these results does not allow to discuss deviations between the scales by NPL and Apel-Riemer Inc. Corresponding additions have been added to the results/calibration section. "The systems FZJ_B, FZJ_A, MHD and PUY, used different certified NMHC calibration standards (Table 4b). If a systematic offset between different scales exists, it should result in systematic deviations from the reference values. MHD, FZJ_B and FZJ_A all used calibration standards from Apel Riemer but the observed deviations from the reference values are random (e.g. deviations for alkanes are of different extent and sign (Figure 2n,q and w). Obviously other instrumental issues (e.g. chromatographic resolution, non-linearity of MS-detector) affected these results and therefore systematic differences between the different calibration scales cannot be assessed. "

ii) Subchapter 2.6 describes the procedure used to estimate the systematic uncertainty for the measurements. While this is useful for an assessment of the actual measurement uncertainties, there are some problems connecting this error analysis to the measurement performance presented in Tables 6 and 7. As mentioned above, uncertainty in calibration standards will not show in a comparison of measurements using the same standard.

The reviewer is right, however, this is not the intention of Table 6-7 (now moved to Supplemental material). In these tables the conformity with GAW and ACTRIS DQOs is reported. Only the colours yellow and red contain information about the uncertainties. For these large deviations the systematic uncertainty contribution due to the standard

C4651

is negligible. The certified uncertainties of the NPL mixtures are at 2% and deviations discussed here are typically larger than this, e.g. beyond the DQO of 5%. In this paper it is of interest to analyse whether deviations of larger than 5% are within the estimated uncertainty. For deviation >5% the systematic 2% calibration gas uncertainty is minor and not considered.

For measurements using the NPL standard therefore a different uncertainty estimate would be required for comparison between uncertainty estimates and results in Tables 6 and 7.

For the three laboratories that determined the composition of the reference sample any truly systematic error most likely will bias the results of the determination of reference sample composition as well as the sample analysis. Since only three laboratories were involved in analysis of the reference samples, bias common to reference value and analysis of the reference sample may be substantial.

As discussed before, the bias is in the range of the certified uncertainty of the NPL scale of 2% is considered but not substantial. The statistical problem of only three reference laboratories remains, however, the median of the results by all other participants are consistent with the reference values. 72% of all median values deviate less than $\pm 5\%$ (ACTRIS DQOs) and only 16% of all medians differ more than 10% from the reference values.

These three different categories of data need to be distinguished more clearly, both in the presentation and discussion of the results. Combined with the fact that only some of the laboratories provided a complete error analysis I am also not sure how to interpret Figure 5, although I think the conclusion "An important result of this intercomparison is that most participants substantially underestimated their uncertainties..." not very surprising.

iii) Not all results presented here meet the criteria of a blind intercomparison.

C4652

We stated more clearly that this is no formal blind intercomparison. It is stated (page 10431, lines 18-23) how EMPA and HPB took part. We further clarified that for Empa and DWD the composition and mole fractions were not completely unknown as for the other participants (see also comment to reviewer 1).

“The composition and the mole fractions in the cylinders were unknown to all participants, except for the reference laboratories HPB and Empa (see above).”

The three laboratories involved in the analysis of the reference samples must have known the reference values at least to some extent. Furthermore, all laboratories must have been aware that they were conducting repeat analysis of the same samples when determining repeatability.

This is according to the definition of repeatability (see GAW glossary “repeatability condition: condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time”, <http://gaw.empa.ch/glossary/glossary.html#2.20>) which is the standard deviation of repeated measurements under the same conditions.

iv) From Tables 6 and 9 it is obvious that many of the laboratories only reported results for a small subset of the ACTRIS target compounds in the air sample. Consequently the summarizing statements such as “Considering the complexity of the matrix and the partly low mole fractions, this is a good result” We decided to skip this sentence. The reviewer is right and only parts of the problems and limitations can be addressed here.

v) Since only one air sample with, compared to background air, generally relatively high molar fractions was used in this intercomparison, it is not obvious to which extent the findings presented here can be used to evaluate the performance of VOC instruments at remote locations. A comparison of the molar fractions of VOC in NMHC_{air} with the range of VOC molar fractions reported for already active sites would help the reader to better understand this problem.

C4653

We added in the method section the following sentence: “Compared to previous intercomparisons, this study used NMHC_{air} with comparable or lower mole fractions (Plass-Dülmer et al., 2006). However, remote stations frequently are exposed to substantially cleaner air (Helmig, 1997; Helmig et al., 2008; Read et al., 2009).”

The authors are aware that the air test gases are typical for rural European air. In line with the air composition characteristic at most of the participating stations testing the performance of real background conditions like encountered at PICO or Cape Verde (REF) were not the focus of this round robin intercomparison.

vi) Most of the discussion deals with details of the potential problems of VOC measurement techniques that are of limited interest and relevance for most users of VOC data. Moreover, the various types of VOC measurement problems that are identified and discussed are well known and have already been described and explained in numerous publications. The finding that column degradation, peak overlays, blank values, incomplete trapping, incomplete desorption or insufficient deactivation of transfer lines will adversely impact reproducibility and accuracy of VOC measurements is nothing new. Similarly, the steps taken to identify the specific problem are based on standard, well established “trouble shooting” techniques in gas chromatography and VOC analysis.

The reviewer is right in his comment however, it reflects the state of the art: The authors tried to, and think it is important to demonstrate the attempts made to understand and explain the reasons for the deviations. We consider this is appropriate for AMT.

vii) The usefulness and value of the paper in its current form is greatly reduced by the limited number of readers who will be willing to read a 63 page paper in order to obtain some straightforward information such as repeatability and difference between a reference value and measurement for a VOC monitoring network. The paper needs to be shortened substantially (maybe to less than 20 pages in total) presenting the gist of the most important results. The very detailed presentation and discussion of

C4654

the results should be moved to the supporting material, something that to some extent already has been done for measurements by proton transfer mass spectrometry. Frequency distributions of the observed deviation from target values or reproducibility are a simple way to present a quantitative, compact and easily digested overview. Such frequency distributions could be presented for all reported data but also include differentiation based on groups of compounds (such as alkenes, light alkanes, aromatic VOC), types of instruments, use of different standards or even measurements impacted by specific problems such as peak overlay or blank values. This would provide an objective overview to the reader. On a positive note, the paper demonstrates that state of the art techniques exist that allow VOC measurements in monitoring networks with overall good quality. This is an important finding, although the somewhat extreme focus on GAW and ACTRIS quality objectives tends to obscure this a bit. I am positive that a substantially shortened version (with details moved to supporting material) will find many interested readers who actually may read more than the abstract.

The authors decided to shorten the paper and to place several Tables and Figures in the supplementary material. Furthermore, Table 1 and Table 5 were merged to one. We added frequency distribution plots which were an excellent suggestion.

Some details to address: i) Precision, as defined in Equation (7) is different from the conventional definition of precision in Analytical Chemistry and also differs from the relative standard deviation of repeat measurements (repeatability) which is presented in Tables 8 and 9. The rationale for the use of this somewhat unconventional definition (including the linear propagation of errors) needs an explanation.

For very small mole fractions the random error is not sufficiently determined by the repeatability but additionally by the detection limit (Equation 7). We refer to the glossary of GAW where it states: "measurement precision - closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions . . . Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation,

C4655

variance, or coefficient of variation under the specified conditions of measurement. . . Measurement precision is a measure of the dispersion of values. . . Precision depends only on the distribution of random errors and does not relate to the "true" value or to the specified value." (<http://gaw.empa.ch/glossary/glossary.html#2.20>). Accordingly, we combined the random errors of VOC measurements into this quantity. In Table 9 (now Tables S3 and 4) the repeatability and not the precision was compared in accordance to data DQOs (Table 2).

ii) There are several compounds for which many laboratories report results for the artificial NMHC sample, but not for NMHC in air. Is this a consequence of problems with detection (or quantification) limits?

There are various reasons for this including detection limit and peak overlays but decisions on data submission were in the responsibility of the participants.

iii) There is no explanation how the integration error (δA , Equation 10) is determined. It also needs to be explained why δA_{sample} is considered as systematic error and not as contribution to the random error determining precision.

Explanations and the derivation of Equation 10 is presented in the Supplementary Materials. See also answers to Reviewer 1.

iv) The headings of Tables 6 and 7 require clarification of what is presented. I assume it is the relative difference between target value and reported value in %. I assume that negative values denote an underestimate for the reported molar fractions? This has been more clarified by: "Overview of NMHC measurement performance in NMHC_N2: relative difference [%] between measured value and assigned value (positive value: higher value than assigned; negative value: lower value than assigned)".

v) Table 9: It looks to me that the values for compounds with molar fractions $<0.1 \text{ nmol mol}^{-1}$ are reported not as relative repeatability in %, but as molar fractions, clarify.

The reviewer is right and this was now clearly stated in the caption of Table 9 (now

C4656

Table S4). "For compounds with mole fractions < 0.1 nmol/mol the repeatability values are given in nmol/mol."

vi) The authors should pay attention to significant digits. Many values are reported with 3 or 4 significant digits, although repeatability and errors in many cases would only justify one or two significant digits. Presenting a measured molar fraction with 4 significant digits suggests an unrealistic accuracy. Using more significant digits than justified may be useful when exchanging data to avoid rounding errors propagating into calculations based on these data. However, the values presented here are final results and should be rounded correctly. See comment above. In case of mole fraction below 0.1 nmol/mol the differences to the assigned values in nmol/mol are reported and not a percentage value. The authors clarified this point in all Tables. Just Table 1 gives 3 significant digits of ~1 ppb values with stated uncertainties of typically below 10 to 20 ppt.

vii) In Table 7 for compounds with <0.1 nmol mol⁻¹ the difference between target value and reported value is given as percentage, in Table 9 as molar fraction. The rationale underlying this difference in type of reporting is not obvious and needs to be explained. We clarified this point in all Tables (see comments above). viii) Table 5 includes molar fractions of VOC for which no results are presented. This is unnecessary. Results for all compounds listed in Table 5 (now Table 1) are presented in the NMHC_air results tables.

Helmig, D.: Ozone removal techniques in the sampling of atmospheric volatile organic trace gases, *Atmospheric Environment*, 31, 3635-3651, 1997. Helmig, D., Tanner, D. M., Honrath, R. E., Owen, R. C., and Parrish, D. D.: Nonmethane hydrocarbons at Pico Mountain, Azores: 1. Oxidation chemistry in the North Atlantic region, *J. Geophys. Res. D Atmos.*, 113, 2008. Read, K. A., Lee, J. D., Lewis, A. C., Moller, S. J., Mendes, L., and Carpenter, L. J.: Intra-annual cycles of NMVOC in the tropical marine boundary layer and their use for interpreting seasonal variability in CO, *J. Geophys. Res. D Atmos.*, 114, 2009.

C4657

Interactive comment on *Atmos. Meas. Tech. Discuss.*, 7, 10423, 2014.

C4658