

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.

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 interest in VOC monitoring data for scientific and regulatory purposes. Understand-

C3943

ing 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

C3944

determined by this procedure, the true accuracy of the reference sample is not evaluated at all. 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. 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". This can be expected since uncertainty in the calibration standard will cancel when comparing measurements using the same calibration standard. For measurements based on a calibration standard that is independent of the NPL standard, uncertainties in both calibrations will propagate into 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.

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

C3945

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. 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 find 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. 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.

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" or "77% of the reported values were within the GAW DQOs, but only 48% were within the ACTRIS DQOs" do not completely convey the existing problems and limitations.

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. vi) Most of the discussion deals with details

C3946

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.

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 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

C3947

version (with details moved to supporting material) will find many interested readers who actually may read more than the abstract.

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.

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?

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.

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?

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.

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.

C3948

vii) In Table 7 for compounds with $<0.1 \text{ nmol mol}^{-1}$ 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.

viii) Table 5 includes molar fractions of VOC for which no results are presented. This is unnecessary.

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