

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

Anonymous Referee #3, Received and published: 2 December 2014

This article summarizes findings from the measurement of two non-methane hydrocarbon compressed gas standards by a total of 20 European laboratories. Two standards were circulated among laboratories, one synthetic gas mixture and one compressed whole air standard. This study is an impressive accomplishment, which, to the best of my knowledge, surpasses most previous comparable experiments. The experiment was carefully designed and well organized. It's a bit of a pity that there were no par-

C4659

ticipants from outside of Europe, despite the fact that the article seems to be guided by Global Atmospheric Watch program and quality objectives. This apparently is due to the funding constraints of this study, which unfortunately only provided support to European partners.

In fact, this study was financed by ACTRIS (European Infrastructure Network) and therefore no participants outside Europe took part in this intercomparison. During the kick-off meeting of ACTRIS the GAW data quality objectives (DQOs) were revised and new DQOs (as described here with ACTRIS DQOs) were formulated. The ACTRIS DQOs will be implemented as the “new” GAW DQOs (GAW VOC-SAG Meeting, Daejong, South Korea, 2014).

The manuscript is very detail oriented providing experimental variables and results for each of the participating laboratories. While I consider this study highly valuable and would like to see these results eventually to be published, I find the manuscript to have a number of weaknesses and to be misleading in its presentation.

The discussion of the ‘performance’ of laboratory centers around two quality objectives, those that were defined for the GAW program and a second, more stringent standard as defined by ACTRIS. From reading the cited GAW document it is my understanding that these target measurement uncertainty windows are for analysis of NMHC in ambient air. The study described in this manuscript, however, did not include any ambient air sampling, but builds on the analysis of two compressed gas cylinder standards. This strategy, of course, makes sense given the much more demanding logistical hurdles for an in-situ intercomparison. My criticism primarily is that findings are presented and discussed in a manner as if these results reflect performance for measurements in ambient air. But this certainly is not the case. This experiment has many limitations due to its artificial approach.

It has not been stated at any place in the manuscript that these are results for real ambient air with the corresponding mole fraction of ozone, humidity and other, potentially

C4660

interfering compounds. To us, it was clear from a concept of such a round robin type intercomparison exercise, that no ambient air is directly tested. Furthermore, previous hydrocarbon intercomparison experiments (NOMHICE; AMOHA, GAW) also used compressed cylinders or flasks as test gases. Accordingly, this difference was not further discussed. However, from the criticism of the reviewer, we realize this potential misunderstanding and included corresponding statements into the revised manuscript. For example we included comments to ozone and water content in the sections preparation of the mixtures, alkene artefacts and other issues.

Mole fractions in the two standards that were used, for most of the compounds, are higher than what typically would be observed at GAW background stations. As also Reviewer 2 had a comment to this, we added in the method section the following sentence: "Compared to previous intercomparisons, this study used NMHC_{air} with comparable (Plass-Dülmer et al., 2006) or lower mole fractions. However, remote stations frequently are exposed to substantially cleaner air (Helmig, 1997; Helmig et al., 2008; Read et al., 2009)."

Analysis of these standards does not consider the challenges of managing interferences from ozone, water vapor, and carbon dioxide, which are present in ambient air at much, much higher levels than NMHC. Many of the participating laboratories didn't even use an ozone management method (Table 4), which is rather surprising and concerning given that ozone scrubbing techniques have become a standard procedure in ambient VOC monitoring. The poorer performances seen in the analysis of the real air standard (compared to the synthetic air) is a first good indication about the added complexity in the analysis of real air. One would expect significantly further degraded agreement with ozone and water vapor present in the sample. We agree with the reviewer in all listed points except for carbon dioxide which is present in NKHC_{air} in similar quantities as in ambient air. However, with respect to the also mentioned high logistic demands for an in-situ comparison, we could not afford to have a side-by-side NMHC intercomparison.

C4661

There have been several other prior extensive experiments of this kind, dating back 20 years, and all relying on the exchange of compressed synthetic or real air standards, including the Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE) [Apel et al., 1994; Apel et al., 1999; Apel et al., 2003], Accurate Measurements of Hydrocarbons in the Atmosphere (AMOHA) [Slemr et al., 2002] [Plass-Dulmer et al., 2006] and a Global Atmospheric Watch coordinated study [Rappenglueck et al., 2006]. Publications that have resulted from this work have pointed out similar findings and directed the community in improving NMHC analysis procedures. For instance, the last publication of the NOMHICE project, published in 2003, presents these primary recommendations "(1) National Institute of Standards and Technology (NIST) standards or NIST-traceable standards should be used and, for mass spectrometric analyses, multicomponent NIST-traceable standards should be used; (2) if solid adsorbents are used for preconcentrating NMHCs, extensive tests should be performed to test for artifact formation and compound losses; and (3) for whole air sampling in canisters, subsequent analyses should be performed as soon as is reasonably possible to avoid the potential for compositional changes." This new publication describing the ACTRIS study only marginally touches on the findings and recommendations that resulted from these prior large experiments. I would find it desirable to show more clearly how ACTRIS built on NOMHICE, AMOHA, and the GAW experiment, how results compare, what new progress has been made since this earlier work, and what newly identified needs and directions for the future are. One obvious change is the move towards adsorbent-based pre-focusing techniques. While more than half of the groups in [Apel et al., 2003] used glass beads/cryogenic focusing, methods listed in the ACTRIS study exclusively relied on solid adsorbent trapping. This is a major shift in technology. Unfortunately, this new paper misses out on the opportunity to thoroughly discuss and evaluate this transition. The previous studies repeatedly stress the importance of utilizing NIST- or NPL-traceable standard scales. It would be desirable for the ACTRIS manuscript to pick up on this topic and provide a more in depth evaluation of how results have improved as a function of the type of standards that were available

C4662

and used by individual laboratories and the standard age. Basically, it was the ambition of the authors to go into such issues as raised by the reviewer and we regret that, as demonstrated in the paper, many of the reported results were not of sufficient quality to evaluate different concepts of, e.g. pre-concentration, standards used, column effects etc. In most cases, no unambiguous identification of superior methods or materials was possible because individual results pointed into different directions than the majority of the other results, probably for other reasons but this could not be resolved. Of the raised points with respect to NOMHICE, the current status is that all laboratories used multi-component mixtures (1) traceable to National Metrologic Institutes like NPL or NIST, thus, an improvement compared to NOMHICE is expected but cannot be discerned from results of laboratories without such traceable standards – such labs did not participate. Also, except for HPB all laboratories used adsorbent trapping (2) and again the difference cannot be proved. However, it is clearly stated that good results can be achieved by laboratories with adsorbent traps. Point (3), canister sampling, was not involved and cannot be commented on. An improvement compared to previous intercomparison studies is hard to determine because different test gases and calibration gases were used. One of the central conclusions of the current study is “Essential for high quality results are experienced operators, comprehensive quality assurance and quality control, well characterised systems, and sufficient man-power to operate the systems and evaluate the data”, or in turn: insufficient experienced operators, QA/QC and manpower are still the major problems and prevent a look onto more technical issues. This is in line with the concepts of ACTRIS and GAW that it is essential to provide standardised QA/QC procedures in the framework of comprehensive measurement guidelines, which are currently developed by ACTRIS and GAW. The results of this round robin intercomparison were expected to demonstrate the status of current VOC measurement capabilities 10-20 years after AMOHA, GAW and NOMHICE inter-comparisons and express the need for better data that will be used in the framework of climate change in order to validate emission inventories and the success of reduction strategies for various gases including VOC.

C4663

The two test mixtures that were circulated were measured prior and after the circulation by three of the participating laboratories. It appears that those laboratories exchanged results while other laboratories did not know the standard compositions. This obviously results in two groups of participating laboratories with much different starting conditions. Surprisingly, statistics of results are not differentiated in that way. I think it would be better to separate laboratories depending on how they participated and which data resulted from a blind robin comparison manner and which did not.

The authors wrote explicitly in the text that HPB and Empa had information about the mixtures and all other had no information about this. However, this was even more clarified in the revised manuscript. See Sections determination of assigned... (line 183) measurement approach (line 233), MS Systems (line 538)

I find the organization of the results discussion difficult to follow. I imagine that for most readers it is not of that much interest to learn how each individual lab performed for each of the standards and for individual compounds. Instead, I think it would be preferable to significantly shorten and reorganize the article about specific analytical topics, for instance enrichment system performance, water management, CO₂ effects, column effects, detection, etc., and provide a summary discussion of what was learned.

I think there is a fundamental misunderstanding on the definition and use of the carbon response factor. As far as I am aware, in the early literature, the flame ionization detector response factor was introduced as a variable describing the dependency of the voltage signal of the detector on the number and bonding type of carbon atoms in the molecule detected. In this paper here, however, this term is being used to describe the performance of the enrichment system, injection, separation, and detection, including possible co-elution of compounds procedure and maybe even regulator effects and the storage behavior of compounds in the compressed air tanks. The detector response is only one variable contributing to the overall 'system response factor'. There certainly is interest in determining this variable, but I think it is not correct to treat and compare it with other studies that evaluated FID response in a more isolated fashion. This ap-

C4664

pears to be a misunderstanding of the intended use of the C-response concept in this study. We fully agree with the reviewer that the original use was to understand the FID detector response. However, we built on this well established response behavior and interpret deviations from an expected response as problems in the respective systems of the intercomparison participants. The big advantage is that one can identify losses (reduced C-response of the entire system) which must be due to parts of the analytical system, e.g. breakthrough in pre-concentration or adsorptive losses to surface materials. This builds on the concept developed by Plass-Duelmer et al. (2002).

It seems odd that discussion of the PTR-MS results is condensed to a mere four lines while results from other labs and systems are presented in much more depth. It was not the scope of this manuscript to compare GC-techniques to PTR-MS. But as two PTR-systems measured these tanks we added these results. In the supplementary material we considered as much as possible the PTR-MS results.

Attention should be paid to the question if sample volumes of calibration standards and of samples withdrawn from the two Round Robin cylinders were the same or different, and if (in case volumes were different) this effected results. Most stations had the same volumes for calibration and test gas samples.

The article mentions several times that there is only one commercial sample pre- focusing instrument in use (Perkin Elmer), but then, later on, two further instrument manufacturers are mentioned (Markes and Entech). This is a misunderstanding. Perkin Elmer is the only manufacturer which sells an entire analytical system (preconcentration unit and GC-system). Markes and Entech are manufacturers of adsorption-desorption systems, but they don't offer complete GC-system.

Please provide more detail on the sample injection methods used. Was air removed from the focusing trap prior to desorption? Was the trap constantly purged during desorption or was there a pre-injection flash heat period? Did system performances show any dependence on the type of carrier gas used?

C4665

All traps were purged prior to desorption to remove air and also during desorption. Also all traps were purged during desorption. All laboratories used helium as carrier gas. This information is now added to the revised manuscript in the header of Table 4.

Benzene and toluene appear to be more of a challenge to quantify in ambient air due to the larger blank values, in particular during real ambient air sampling when using solid adsorbents. It would be valuable to discuss this issue in more depth in the light of the findings from this study. Air measurements at remote stations all face the problem of correct blank value consideration (for all compounds). We discussed the importance of good blank characterization related to alkene measurements which revealed on average higher blanks than aromatics. Also, the conclusion states that "The study highlights the importance of good zero-gas measurements to determine realistic blank values to be subtracted from measurement results."

In summary, I recommend that this manuscript be re-organized, significantly shortened, and that findings be clearly presented as an artificial laboratory intercomparison, omitting critical and very important variables that need to be considered for real world ambient air measurements at background locations, and that consequently, data and evaluations presented here, only reflect a part, but not the whole procedure for NMHC analysis in ambient air. In that sense, this manuscript, unfortunately, does not present much progress and novelty in the approach of the experiment (no real ambient air sampling) and in the data interpretation compared to earlier, in many cases more than ten years old literature. In the last 10 year no new publication on intercomparisons were published (please provide us newer Round Robin publications).

We followed the suggestions and have considerably shortened and included corresponding text into the revised manuscript. Tables 1 and 5 were merged to a new one (Table 1 in the revised version) Tables 6-10 in supplementary material Figures 2-3, 6 in supplementary material New Figure 3 (frequency distribution plot) to replace Tables 6-7 and Figures 2-3 Text was considerable shortened mainly in the result and discussion part

C4666

Minor comments: 10425/1-3: Mention that this project addresses analysis in ambient air.

The authors changed the sentence to: "The performance of 18 European institutions involved in long-term non-methane hydrocarbon (NMHC) measurements in ambient air within the framework. . ."

10425/24:....breakthrough of C2-C3 hydrocarbons in the focusing trap ... Changed as suggested.

10425/27: Besides what is mentioned here, as previous studies have emphasized, a well-established and cross-referenced calibration gas scale is also imperative for high quality results. This was understood as one part of "comprehensive quality assurance and quality control" but we added this explicitly to the abstract.

10426/5: Rather than citing the editor of this series of articles, the particular articles and authors should be cited. Here only a rough overview shall be given and we made clear now that references in these text books should be considered: "(Koppmann (ed.) , 2007, and Warneck, 1988, and references therein).

10426/19-20: I disagree with this sentence. There are numerous labs in Europe that for a long time have been performing high quality ambient air level VOC measurements. There are several sources for high quality, low uncertainty VOC standards, and audits conducted by the World Calibration Centre for VOC has proven that many labs do a very reasonable job in their analysis. We changed this into: "The measurement capability for VOCs in Europe is widespread. Though several long-term and high-quality laboratories exist that are building on sophisticated QA/QC, high quality standard gases, previous intercomparison activities, and audits by World Calibration Centre for VOC, other laboratories lack from the fact that there are no commonly agreed, homogenized quality assurance procedures, and measurement guidelines ."

10427: I suggest to not citing urls in the text. The authors think that it is important to

C4667

provide the ACTRIS link, as this framework financed this intercomparison.

10427/9: Please explain 'coverage factor'. This roughly corresponds to a 95% confidence interval and this is generally used in gas metrics (e.g. NMI references, climate gases references, GAW glossary, VOC Measurement Guidelines). An explanation was added. See also comment to Reviewer 1.

10429/11: ...and one with NMHC_air. Changed as suggested.

10430/1:The mole fractions of the C2-Cx NMHC in NMHC_air ... Changed as suggested.

10432/17:....trimethylbenzenes and ... Changed as suggested.

10433/2:.... .and they were envisioned to be implemented as new ... Changed as suggested.

10432/5: I would call these DQO as 'goals', rather than as a level of 'acceptance'. The sentence was adapted as follows: "The goals in repeatability of the measurements in ACTRIS are 2% for alkanes, alkenes (including isoprene), alkynes, and aromatics, and 5% for monoterpenes. For the uncertainty, the goals for deviation from a reference value are set to 5% for alkanes, alkenes (including isoprene), alkynes, and aromatics, and to 10% for monoterpenes."

10437: I find 'non-compliant' to be a rather poor word choice here. I guess what is meant is 'deviations larger than the stated quality objectives'. Changed as suggested.

10437/22: Make clear if '10 %' is relative or absolute. We intended the relative repeatability. We clarified the sentence and added the word "relative".

10440/4: I don't think this statement is accurate. There are a series of variables that effect the FID response and those variables seem to vary somewhat between instrument brands and individual machines. In this study, no significant differences could be attributed to structural elements of VOC compounds, like alkenes, alkanes, or aro-

C4668

omatics in either manufacturer's FID type (Figure 4). Instead, deviations of individual compounds from their homologous series, e.g. hexane from other alkanes, were observed. Such deviations can hardly be understood as being due to the respective FID and its operating conditions.

10440/20-21: Poorly worded. Better say "...there could be several reasons for..." We rephrased the sentence as follows: "In general, there are several potential reasons if the normalized for the deviations of C-response factors from between the calibration standard and the NMHC_N2 differ from each other (Figure 4) or and from the expected value of 1."

10442/11: Ethyne is a C2 hydrocarbon, so shouldn't it be included in section 3.5? Or maybe section 3.5 could be broken up into 3.5.1 and 3.5.2?

As ethyne does not behave like other C2 compounds, the authors decided to discuss this compound separately.

10442/16: The ethyne ECN is based on very few and quite old work. It is obviously not well defined, probably also varying significantly between instruments used, which would be a good point to make here. We added a corresponding remark: "... the effective carbon number is between 2 and 2.6 indicating higher uncertainty of the C-response for this compound." However, we don't expect old studies to be less accurate.

10457: Last row of Table 2: move "< 0.1 . . ." up to row above. Changed as suggested.

10459: Provide particular adsorbent specifics for 'Air Toxics Analyzer'. Provide the mass of adsorbent used in the pre-focusing traps? The authors asked the manufacturer, but as this is a trade secret, we don't have further specifications.

10461: Why were two Nafion Dryers used in some of the systems? This is because the Medusa systems need two Nafion dryers. Between trap 1 and trap 2 a second Nafion dryer is needed to remove humidity from the sample, which was preconcentrated on trap 1 and the humidity was not completely removed from the first Nafion. For further

C4669

details see Miller et al. (2008). (Miller, B. R., et al. (2008), Medusa: A sample preconcentration and GC/MS detector system for in situ measurements of atmospheric trace halo-carbons, hydrocarbons, and sulfur compounds, *Anal. Chem.*, 80 (5),1536–1545, doi:10.1021/ac702084k.)

10462: Add information on when calibration standards were prepared. We added this information to the paper in Table 4 last column.

10467-10471: The font size used in these tables is too small to be readable in a printout hardcopy, which in my opinion is unacceptable. Turning these tables (converting rows to columns and vice versa) would allow for more space and larger fonts. The authors tried this already. We will talk to the setting specialist of Copernicus if they can change something. However, as this is an online Journal, the tables can be printed in another size to have it better readable.

10475: As mentioned above, as far as I can read the very small print, these data do not reflect the 'C-response of FIDS', but instead show the relative recoveries of the entire sample introduction and analysis procedure. We change the table caption in to "FID-C-responses as indicator of NMHC measuring system artefacts."

10477-10480: Font size is too small. The authors tried this already. We will talk to the setting specialist of Copernicus if they can change the font.

References cited Apel, E. C., J. G. Calvert, and F. C. Fehsenfeld (1994), The Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE) - TASK-1 AND TASK-2, *J. Geophys. Res.- Atmos.*, 99(D8), 16651-16664. Apel, E. C., J. G. Calvert, T. M. Gilpin, F. Fehsenfeld, and W. A. Lonneman (2003), Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE): Task 4, ambient air (vol 108 art no 4359, 2003), *J. Geophys. Res.-Atmos.*, 108(D12), doi:10.1029/2003jd003783. Apel, E. C., J. G. Calvert, T. M. Gilpin, F. C. Fehsenfeld, D. D. Parrish, and W. A. Lonneman (1999), The Nonmethane Hydrocarbon Intercomparison Experiment (NOMHICE): Task 3, *J. Geophys. Res.-Atmos.*, 104(D21), 26069-26086. Plass-Dulmer, C., N. Schmidbauer, J.

C4670

Slemr, F. Slemr, and H. D'Souza (2006), European hydrocarbon intercomparison experiment AMOHA part 4: Canister sampling of ambient air, *J. Geophys. Res.-Atmos.*, 111(D4), doi:10.1029/2005jd006351. Rappenglueck, B., et al. (2006), The first VOC intercomparison exercise within the Global Atmosphere Watch (GAW), *Atmospheric Environment*, 40(39), 7508-7527, doi:10.1016/j.atmosenv.2006.07.016. Slemr, J., F. Slemr, R. Partridge, H. D'Souza, and N. Schmidbauer (2002), Accurate Measurements of Hydrocarbons in the Atmosphere (AMOHA): Three European intercomparisons, *J. Geophys. Res.-Atmos.*, 107(D19), doi:10.1029/2001jd001357. Interactive comment on *Atmos. Meas. Tech. Discuss.*, 7, 10423, 2014.

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

C4671