

## ***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.***

**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 petty that there were no participants 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

C4002

European partners.

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. 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. 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.

There have been several other prior extensive experiments of this kind, dating back ~ 8 – 20 years, and all relying on the exchange of compressed synthetic or real

C4003

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 and used by individual laboratories and the standard age.

The two test mixtures that were circulated were measured prior and after the circulation  
C4004

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.

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 re-organize the article about specific analytical topics, for instance enrichment system performance, water management, CO<sub>2</sub> 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.

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.

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.

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).

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?

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.

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.

Minor comments:

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

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

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

C4006

quality results.

10426/5: Rather than citing the editor of this series of articles, the particular articles and authors should be cited.

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.

10427: I suggest to not citing urls in the text.

10427/9: Please explain 'coverage factor'.

10429/11: ... and one with NMHC\_air.

10430/1: ...The mole fractions of the C2-Cx NMHC in NMHC\_air ...

10432/17: ...trimethylbenzenes and ...

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

10432/5: I would call these DQO as 'goals', rather than as a level of 'acceptance'.

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'.

10437/22: Make clear if '10 %' is relative or absolute.

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.

10440/20-21: Poorly worded. Better say "... there could be several reasons for ..."

10442/11: Ethyne is a C2 hydrocarbon, so shouldn't it be included in section 3.5? Or,

C4007

maybe section 3.5 could be broken up into 3.5.1 and 3.5.2?

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.

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

10459: Provide particular adsorbent specifics for 'Air Toxics Analyzer'. Provide the mass of adsorbent used in the pre-focusing traps.

10461: Why were two Nafion Dryers used in some of the systems?

10462: Add information on when calibration standards were prepared.

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.

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.

10477-10480: Font size is too small.

#### 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.

C4008

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. 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 inter-comparisons, *J. Geophys. Res.-Atmos.*, 107(D19), doi:10.1029/2001jd001357.

---

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

C4009