This manuscript, “Towards a high quality in-situ observation network for oxygenated volatile organic compounds (OVOCs) in Europe: transferring traceability to the International System of Units (SI) to the field” describes 2 approaches to create SI-traceable standards for acetaldehyde, acetone, methanol and methyl ethyl ketone in detail. The motivation for this work is described nicely in the introduction. There are three scientific goals of the paper: to create reference gas mixtures (RGMs) in dry nitrogen at amount fractions of ~100 nmol mol^{-1} that are stable in time, to create working standards at amount fractions on the order of 10 nmol mol^{-1} by diluting the RGMs, and to create working standards at amount fractions on the order of 10 nmol mol^{-1} by spiking whole air samples. This work is highly collaborative and detailed. The sections are well-organized.

General Comments:

1. In general, the paper is highly detailed. I think a schematic describing the overarching process for creating these working standards could assist with the clarity of the paper, and anywhere a schematic could reduce details included as text would be appreciated (i.e. for the filling of the parent and subsample cylinders, or the diluting system).

   Following the valuable recommendations of referees 1, 2 and 3, we will add three schematics illustrating the different processes involved in this work: from generation or preparation of the SI-traceable working standards to their assessment (see answers to referee 2).

2. It is difficult to compare the two methods used to create working standards. For diluted RGMs, the results are shown in percent difference compared to the assigned amount fractions, and for the spiked whole air samples the results are shown in amount fractions plotted alongside the assigned value. Do the authors have a reason for providing the results in this way? It may be more helpful to future users of these techniques to present both in units of percent differences.

   We thank the referee for the comment. The reason for providing the results in that way was to show the low amount of fraction levels measured as well as the difference between the spiked and the certified OVOC amount fractions. To help readers and users, we changed y-axis of the figures for the SI-traceable working standards based on certified spiked whole air samples to units of percent differences.

3. In section 4.1, the authors describe the methods used to assess the working standards based on whole air samples. My understanding is that the cylinders were large enough to be sent to all the laboratories, but canisters only had enough volume “for one analysis”. Could the authors comment on their recommendations for going forward with this technique? Would monitoring stations be expected to purchase a cylinder (for long term use) or a canister (for one time use) of working standard? Does the cost of manufacturing these standards seem worth the trouble for monitoring stations?

   The referee's understanding is correct. The main limiting factor for the use of canister at monitoring stations is the instability of the whole air sample within the canisters (particularly the Siltek® stainless steel canisters) rather than the small volume of working standard within the canister (i.e., for one time use). Further research is needed but whole air samples within pressurised cylinders seems to be the option for going forward with this technique. The advantage of this kind of SI-traceable working standards is that they will allow monitoring stations calibrate their instruments using standards that use similar matrix gases than their samples (ambient air). Although RGMs are more stable, the matrix gas used is nitrogen, differing therefore from the ambient air. Matrix gas effects on the analytical systems are not fully understood yet, thus these standards might provide some insights on the topic. Although the authors don't have an answer based on solid proofs, the cost of manufacturing and certifying these standards might be worth the trouble for monitoring stations. The discussion was extended to cover this point.

4. Could the authors expand on what they mean by “molar masses” as a source of uncertainty during the preparation of RGMs (line 192)?

   For the assignment of the amount fraction value of the selected OVOCs within the RGMs, the molar mass of each of the selected OVOCs is needed. This molar mass is based on the atomic weights of each of the atoms that form the compound (e.g., the molar mass of methanol is the sum of the atomic weight of carbon (C) and four times the atomic weight of hydrogen (H)). Atomic weight values are provided with an uncertainty (Coplen et al., 2020; van der Veen et al., 2021) that needs to be considered in the molar mass of each OVOC. Because, as said, the molar mass is part of the equation used to estimate the amount fraction of the OVOCs and it has an uncertainty associated propagated from the atomic weights, molar mass is considered a source of uncertainty of the prepared RGMs. We added these references to the manuscript.

   Coplen, T.B., Holden, N.E., Ding, T., Meijer, H.A.J., Vogl, J., Zhu, X. (2020). The Table of Standard Atomic Weights – An exercise in consensus. *Rap. Commun Mass Spectrom,* 36, e8864. https://doi.org/10.1002/rcm.8864

   van der Veen, A.M.H, Meija, J., Possolo, A., Hibbert, D.B. (2021). Interpretation and use of standard atomic weights (IUPAC Technical Report). *Pure Appl. Chemi.,* 93, 629–646. https://doi.org/10.1515/pac-2017-1002.

5. On line 397 the authors write, “Compound loss after preparation due to surface effects might explain relative differences of around -5 % for ethanol.” Do the authors mean that ethanol had relative differences of around -5 %, and the most likely reason for that is surface effects? On first read, I thought the authors were suggesting they had done a calculation for surface effects and could explain -5 % of the ethanol differences.

   We rephrased the sentence to make clear that the relative differences for ethanol were around -5% and the most likely reason for that is surface effects.

6. On line 430 the authors write, “vessel material may play a role in the lack of homogeneity…”. Was there a material that performed better than others? This is an interesting result.

   Vessel surface material performance was not the focus of this research. Experiments designed specifically to explore effects of the surface material on the stability and homogeneity of the working standards based on certified spiked whole air samples would be needed to be able to answer this question. Because the pressurised cylinders differ on their volume (10 L for cylinder type 1 and 3.6 L for the type 2) and these, in turn, differ from canisters in volume and pressure, it is difficult to state which material performed better. Differences in performance might be due to the other factors (e.g., differences on pressure, surface to volume ratio and cylinder histories) and not to the material. When comparing canisters, with same volume and similar pressures, it seems that Silonite™ stainless steel canisters perform better than Siltek® stainless stee canisters, but this cannot be confirmed because of the different history of the canisters, which might have contributed to these observations. Using new canisters and the right experimental design would allow to draw a conclusion regarding the material with the best performance.

7. On line 549 the authors write, “Assessment results for amount fraction levels around 10 nmol mol^{-1} suggests that SI-traceable working standards based on dilution of RGMs can be used as calibration standard at monitoring stations for key OVOCs, such as acetone. For other OVOCs, like methanol, acetaldehyde and MEK, suitability is not so clear due to the large uncertainty and result dispersion, particularly at amount fractions < 5 nmol mol^{-1}.” However, in the conclusion, line 610 is, “The first type, based on RGMs diluted with dry nitrogen, using for that a portable dilution system that ensures SI-traceability after dilution, seems to be suitable for calibration of acetone, MEK and methanol at amount fractions around 10 nmol mol^{-1}.” These statements seem conflicting as written.

   With the second of the first sentences indicated by the referee, we meant that at lower amount fractions (< 5 nmol mol^{-1}) this type of working standards for MEK and methanol doesn't seem to be suitable. We rewrote this statement to make the information clearer, avoiding the apparent conflict among the statements: "Assessment results for amount fraction levels around 10 nmol mol^{-1} suggest that SI-traceable working standards based on dilution of RGMs can be used as calibration standard at monitoring stations for key OVOCs, such as acetone. For other OVOCs, like acetaldehyde, suitability is not so clear due to the large uncertainty, particularly at amount fractions < 5 nmol mol^{-1}. At this low amount fraction level, suitability of the SI-traceable working standards for MEK and methanol is also questionable".

Technical Notes:

Line 67 - typo “atmosphere and play and important” should be “atmosphere and play **an** important”

Corrected.

Line 68 - “the main sinks are the oxidation with the OH radical and…” would be better worded as “the main sinks are oxidation with OH radicals and…”

Changed.

Line 81 - typo “emphasises” should be “emphasise” or “emphasize”

Corrected.

Line 118 - should “metrological accepted term” be “metrologica**lly** accepted term”?

Thanks for the suggestion. We rephrased as follows to keep the meaning: "the accepted metrological term".

Line 214 - “PTFE” and “SS” are used for the first time, but not defined as acronyms

Definitions of both acronyms added to the text.

Line 233 - typo “consisted on certified” should be “consisted **of** certified”

Corrected.

Line 245 - typo, no need for comma after “both”

Deleted.

Lines 320 and 799 - could not find an Eq. (B1); should a different equation be referenced?

Corrected. Lines 320 and 799 refer to Eq. (C1) and not Eq. (B1).

Line 326 - CRF is used for the first time, but not defined as an acronym

Acronym definition added.

Line 553 - “contributed to **that**”, replace ‘that’ with the true meaning

Done.

Line 671 - typo “the carrier was helium” should be “the carrier **gas** was helium”

Corrected.

Table B1 - the subscripts appear in the order “2,3,1”, consider rearranging to “1,2,3”

Thanks for the suggestion. We rearranged the subscripts accordingly.

The current Table B3 appears in section B2, and the current Table B2 appears in section B3. It seems like renaming the tables (and their text references) would fix this.

Tables were renamed and references in the text corrected (lines 388, 400, 466, 713-714 and 722)

Line 738 - typo, missing a close-parenthesis

Close-parenthesis added.

Line 828 - typo “Prior instrument calibration, …” should be “Prior **to** instrument calibration, …”

Corrected.