The manuscript entitled "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" by Iturrate-Garcia et al. presents the developed and assessed protocols to generate two different types of SI-traceable working standards for acetaldehyde, acetone, methanol, and methyl ethyl ketone. This method enables the calibration of OVOCs at lower concentration levels and ensures the comparability of OVOC measurements both within and across monitoring networks. The manuscript would benefit from clearer articulation, particularly in the explanation of the methodology. Additionally, a more thorough discussion of the method's strengths, limitations, and inherent uncertainties is needed to enhance its scientific rigor. The manuscript may be suitable for publication after major revision.

General comments:

1. The manuscript contains an excessive amount of detailed content, making it difficult for readers to grasp the key points. I recommend improving the readability by incorporating structured steps or flowcharts where applicable. For example, provide a flowchart illustrating the process from standard sample preparation to dilution and measurement.

We thank the referee for the suggestion. We will add three flowcharts illustrating the different processes involved in the assessment of the two different types of SI-working standards to the manuscript.
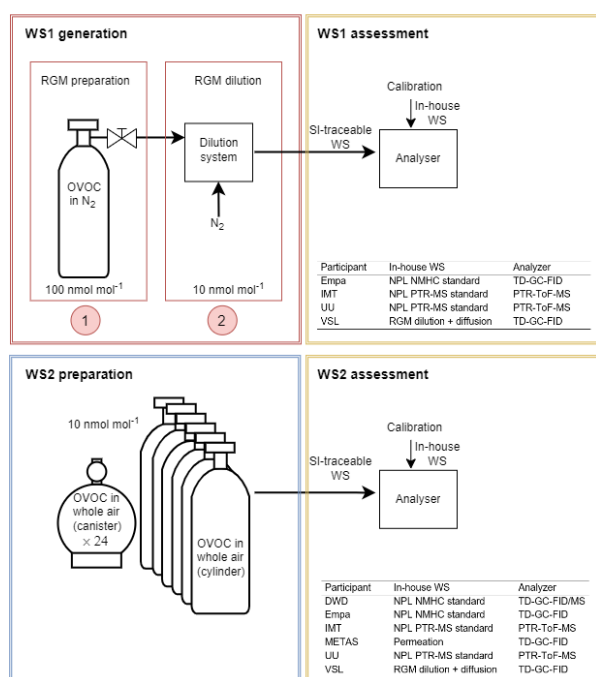
Figure 1: Scheme showing the two types working standards traceable to the international system of units (SI) prepared in this work, based on the dilution of reference gas mixtures (RGM) of oxygenated volatile organic compounds (OVOC) in nitrogen (N_2) (WS1; for details, see section 2.1) and on certified spiked whole air samples (WS2; for details, see section 2.2). Participants in the assessment, analysers (thermal desorption (TD)-gas chromatography (GC)-flame ionization detector (FID) and proton transfer reaction (PTR)-Time of Flight (ToF)-mass spectrometry (MS) systems) and in-house working standards used to calibrate them are indicated.
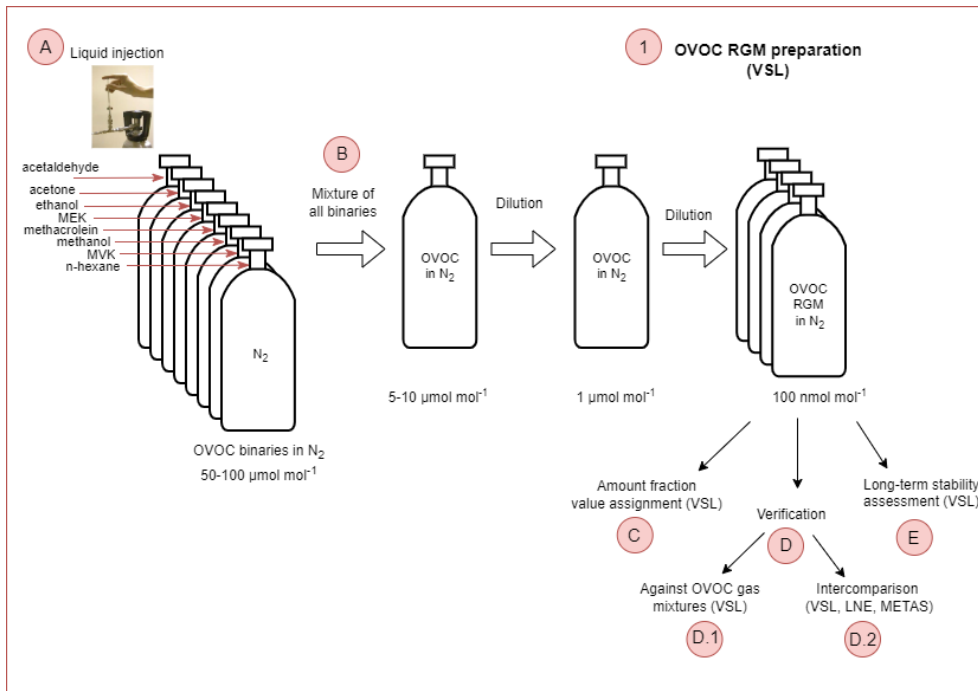
Figure 2: Schematic diagram illustrating the steps needed to prepare the reference gas mixtures (RGMs) of the selected oxygenated volatile organic compounds (OVOCs).
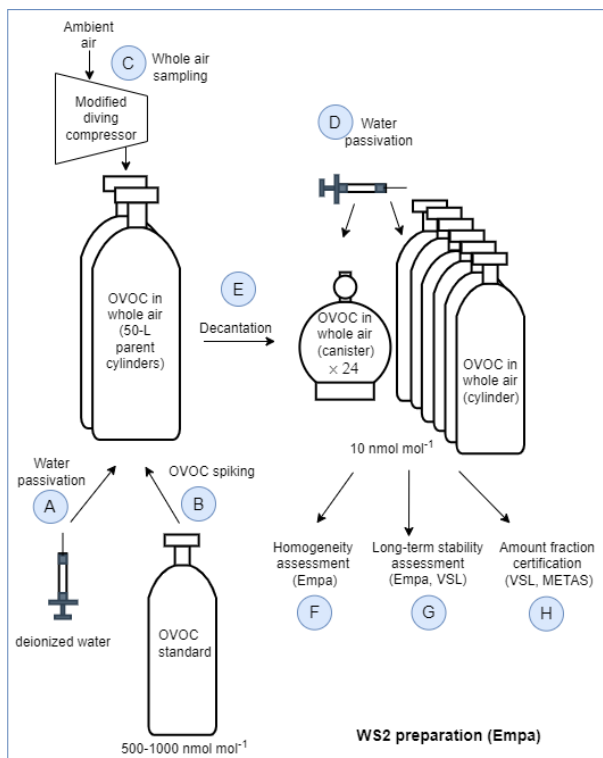
Figure 3: Schematic diagram illustrating the steps needed to prepare the SI-traceable working standard based on certified spiked whole air samples (WS2).

2. Abbreviation Issues:

a) The experiments in the manuscript were conducted across multiple participant laboratories (e.g., VSL, METAS, LNE, IMT, UU, and Empa). The full names of these laboratories should be clearly defined when the abbreviations first appear.

For the National Metrology Institutes, instead of writing the full name the first time the abbreviation appeared, we indicated clearly which national metrology institute (NMI) was the first time the abbreviation appeared in the main text (e.g. NMI of the Netherlands (line 131), NMI of France (line 186), NMI of Switzerland (line 186)), as this information is more relevant than the definition of the abbreviation. For the other participants, the definition was already included in the main text: line 237 (Empa), line 301 (UU, IMT) and line 305 (DWD). To facilitate the reading of the annexes, we added this information the first time the abbreviations appeared in each of the annexes. We did the same for other abbreviations, such as RGM and TD-GC-FID, among others.

b) There are inconsistencies in the capitalization of "Empa" throughout the text. Please ensure uniform usage.

Corrected.

c) Additionally, the description of which laboratories conducted specific experiments using particular methods is somewhat unclear. A brief introduction of this information should be included either at the end of the introduction or at the beginning of Sections 2, 3, and 4 for clarity.

Information on which laboratories conducted specific experiments and the methods used is included in the manuscript annexes. However, the information is split into several tables. For clarity, we implemented the referee's suggestion.

3. The authors provided two different types of SI-traceable working standards, accompanied by two different dilution systems and distinct analytical methods. Although the manuscript outlines the method for calculating uncertainty, it lacks specific uncertainty values for each OVOC corresponding to each method. It is necessary to supplement the manuscript with detailed uncertainty values, the advantages and limitations of each method, and recommendations on preferred methods, processes, and considerations. This will enhance the applicability of the methods within the atmospheric measurement community.

We thank the referee for these recommendations. Regarding the detailed uncertainty values, the uncertainties of the SI-traceable working standards were already included in the manuscript (Table 3 and Table D1). We added the uncertainty of the distinct analytical methods to Table D2 in Appendix D. It is acknowledged that the comment about advantages and limitations could be better formulated in the conclusions. It must also be noted that the use of different analytical techniques (as it is normal practice in the climate community) is also playing a big role on the results of this study and on how the working standards should be used. Concerning the recommendations on preferred methods, with the current results, is difficult to evaluate which of the methods is more suitable for OVOC monitoring, because of the different calibration standards used (not possible to identify if the differences in uncertainties for all the instruments are only due to the calibration standards used or to the analytical system). To be able to give right recommendations about the most suitable method, measurements should be performed using the same calibration standards, for example, using one of the described SI-traceable working standards.

Specific Comments:

1. Line 5: This sentence contains ambiguity: while ozone has a much lower abundance compared to N2 in the atmosphere, its concentration is still relatively high compared to VOCs, and many countries and cities suffer from ozone pollution. It is recommended to remove the statement about low abundance and rewrite the sentence to emphasize ozone's contribution to OH production.

Statement removed and ozone's contribution to OH production emphasized.

2. Line 67: Typo on "play and important role".

Corrected.

3. Line 255: The description and quantity of "cylinders and canisters" in Table C2 differ from those provided in the main text and Appendix C. Please review and verify for consistency.

Corrected. Information on two canisters was missing in Table C2 (MVOC151-005D and MVOC151-008B).

4. Line 388: The tables in the Appendix B should be presented in order in the manuscript.

Corrected. Moreover, the order of the Appendix B subsections was changed, so they are introduced in sequential order in the main text (B.2 RGM verification and B.3 RGM stability evaluation).

5. Line 412-418: In some cases, images can be more vivid and readable than tables. Moreover, the manuscript emphasizes the assessment of four species: acetaldehyde, acetone, methanol, and methyl ethyl ketone (MEK). However, Figure 1 only presents data for acetone. Please include the results for the other species as well.

Figure 1 shows the results of an interlaboratory comparison that was used to complement the RGM verification process based on comparison against OVOC standards. This is not, therefore, part of the assessment of the working standards described in the manuscript. For the comparison, one of the laboratories could only assess three of the OVOCs added to the RGM cylinders: acetone, ethanol and methanol. The results for acetone were shown as an example. In Appendix B (Figure B.1), results for methanol are shown. For ethanol, due to technical issues with one of the analytical systems, the uncertainty of one of the participants (METAS) is too big to be able to drive any conclusion based on these results. Acetaldehyde and MEK was only measured by one laboratory (VSL). Although MVK was measured by two laboratories (VSL and METAS), the low long-term stability of the MVK permeation unit used as reference by METAS resulted in a big uncertainty that prevented us to use them to drive any conclusion.

6. Line 427: Please specify the data source for this paragraph and list it in the Appendix.

The data source for this paragraph is the analysis of variance performed on the data from the homogeneity test. Homogeneity data was included in Appendix C (Table C3).

7. Line 441: In the manuscript, the authors state that "certification results obtained for whole air samples contained in pressurised 10 L aluminium cylinders showed good consistency between the two laboratories performing the certification (i.e., VSL, METAS)." However, there is no data provided to support this claim. Please include the relevant data that demonstrates this consistency and add it to the Appendix for verification.

Data supporting this claim is included in Table C4. We added a reference to this table to facilitate the tracking of these data. We also added a Figure in the Appendix representing the consistency of the measurements between VSL and METAS through the criterion defined in Eq. (4).

8. Line 608 : The conclusions should be drawn cautiously. Based on Figures 2 through 6, it appears that the two different types of SI-traceable working standards may be more suitable for OVOC amount fractions in the range of 4–20 nmol mol^{-1}.

The figures included in line 608 (4–10 nmol mol^{-1}) refer to OVOC ambient levels, which were provided by different monitoring stations during a stakeholder survey performed at the beginning of the EURAMET research project 19ENV06 "MetClimVOC". In Figures 5 and 6, the whole air samples certified and measured are above 10 nmol mol^{-1} (even close to 20 nmol mol^{-1} for acetone), but this is due to the OVOC spiking.