

# Harmonizing measurements of oxidative potential of PM: an interlaboratory comparison of the ascorbic acid assay

## Authors' response

We would like to thank the referees for taking the time to evaluate our manuscript, and for their positive and constructive feedback that will surely improve the quality of our work. Our responses are presented below, in [blue](#).

### Response to referee n#1:

Air pollution is of great concern for the whole humanity for the obvious reasons. In addition to PM<sub>2.5</sub> mass, their oxidative potential (OP) is getting more and more attention in recent years as OP is a metric that potentially links chemical species in aerosols to health effects. This manuscript (MS) summarizes the findings and lessons learnt from an inter-laboratory comparison (ILC) campaign focused on the ascorbic acid assay for OP measurements (OP-AA). The ILC evaluates the capacity of the scientific OP community to implement the protocol and identifies the key parameters that must be strictly controlled to improve reproducibility and comparability. By assessing both intra- and inter-laboratory variability, using a common protocol on identical samples, this work provides valuable insights for the global standardization of OP measurements. Twenty-six laboratories worldwide quantified OP of four PM filter samples and of one chemical compound to evaluate the entire analytical chain, including the extraction step, using a simplified OP-AA protocol. It is a very important study, which is well designed and nicely executed.

I recommend the publication of this MS. However, the following points needs attention for possible betterment and future ILC campaigns.

L134: Give a reference that PAHs has no intrinsic OP

[We now refer to Pietrogrande et al., \(2022\) and Calas et al., \(2017\) to support this statement.](#)

L135: Define “Latent OP” for readers, as it is not a usual terminology.

[“Latent OP” refers to OP attributable to PM metabolization or inflammatory reactions due to non-redox active PM species, while “intrinsic” OP reflects measurable concentrations](#)

of redox-active species or radicals associated with the particle surface. This is explained in the sentence l. 131-135, but we modified the sentence to further clarify this point.

The sentence now reads as: *“Critically, this means that these assays do not capture the additional PM-associated oxidative burden, or ‘latent’ OP, arising from xenobiotic metabolism, or the induction of downstream inflammatory responses, driven by PM constituents that have no intrinsic oxidative potential, such as polycyclic aromatic hydrocarbons and bacterial cell wall components. Therefore, it has been proposed that PM has both ‘intrinsic’ and ‘latent’ OP, with the acellular assays informing the former but not the latter (Kelly et al., 2011a, b).”*

L165-170: It would be relevant to make a case here that why multiple assays are important to assess PM OP. This discussion shall also include limitations of DTT-OP and how AA-OP will reduce the limitations.

Different chemical mechanisms and ROS are responsible for the oxidative capacity of PM in the body, this is why several assays co-exist. In the absence of epidemiological evidence for a more health-relevant metric, several studies have recommended assessing PM OP using at least two assays (Liu and Schauer, 2026, RI-URBANS Guidance document on the measurement of OP in PM - <https://riurbans.eu/wp-content/uploads/2025/10/ST4.pdf>). The purpose of this work is not to defend the thesis that OP-AA would reduce the limitations of OP-DTT, but rather to collaboratively develop a harmonized protocol in the context of measuring OP, using complementary assays. The chemical sensitivities and epidemiological evidence of each probe have been extensively discussed in recent review papers such as the one from Liu and Schauer (2026), this is why these aspects were not further presented in our work.

Nevertheless, we modified lines 173-176 that now read as: *“However, studies using this assay have reported contrasting and inconsistent results regarding health outcomes compared to those using OP-DTT (Liu and Schauer, 2026). Differences in the chemical species and redox pathways captured by each assay, as well as significant variability in experimental protocols may partly explain these discrepancies. Consequently, and in the absence of epidemiological evidence for a more health-relevant metric, the use of complementary assays is recommended, and the harmonization of protocols is urgently needed.”*

L225-227: This step can introduce a considerable uncertainty in the measurements as well as in OP estimation. How the mass of PM used in the measurements was calculated? As the PM extracts were not filtered (as per L211-213), some initially undissolved species may contribute differently because different chemical species may have different

dissolution rates. Some very fine particles can also affect absorbance measurements via physical absorbance of light by these particles. How was this issue addressed?

The mass of introduced PM was calculated based on extract concentration ( $10 \mu\text{g mL}^{-1}$  or  $25 \mu\text{g mL}^{-1}$  depending on the sample) and injected volume ( $80 \mu\text{L}$  and  $1.2 \text{ mL}$  for plate-reader and cuvette protocols, respectively). We agree that this assumes repeatable extraction efficiency, and homogeneous dissolution of chemical species. One previous study has reported a plateau in extraction efficiency for PM samples after 75 min when extracting in ultra-pure water, and after 40 min when extracting in a Gamble + DPPC solution (Calas et al., 2017). This suggests that these durations were sufficient to minimize the sources of variability due to extraction efficiency and dissolution of chemical species for the tested samples, and these performances are very likely applicable to extractions in potassium phosphate buffer. We acknowledge that for specific samples, such choices could cause some variability. By performing three extraction replicates and three analytical replicate per extract, our work aimed at evaluating the impact of such sources of uncertainties in five samples. The variance decomposition showed that for the five samples considered, the repeatability was better than 5% and that the extraction variance accounted for 7 to 14% of the total variance (lines 361-362). This shows that absorbance measurements were only marginally affected by the presence of very fine particles. Sample C, with the highest share of extraction variance in the total variance could be affected by inhomogeneous dissolution of chemical species. We added this hypothesis in the discussion regarding sample C lines 367-369 to clarify this point:

*“Sample C could also have a chemical composition leading to this higher variability, with species that would be less water-soluble, complexed by the phosphate buffer extracting solution, **or with different dissolution rates.**”*

Although repeatability had a minor share of total variance, we also added a sentence in the recommendation section to better detect the cases of light absorbance of particles, lines 698-701: *“To be able to detect absorbance of light by particles, multiple absorbance measurements per replicate could be performed at each time point. High variations between measurements of the same replicate at the same time point would be indicators of such phenomenon, suggesting to either discard the absorbance reads or to reanalyze the sample.”*

Fig. 2: This information can be given in tabular form with greater details including names of institutes/countries, unless participants requested for anonymity. Further, please include a reason for including labs mostly from Europe some from USA, and a few only from other countries. There are many other countries from where OP studies are reported. Including the maximum possible labs with different environmental/

meteorological conditions would have been beneficial for the whole scientific communities.

Thank you for your suggestion and concerns regarding the inclusion criteria. Since most co-authors are participants of the inter-laboratory comparison, their affiliation can easily be found. We believed that a map was more visual than a table. Participating laboratories were not chosen by the organizers, but through a call for participation shared via ACTRIS NF Forum, ACTRIS Community, RI-URBANS task members, and participants to the OP-DTT ILC (see lines 186-189). The inclusion criterion was that the laboratory had to perform the OP-AA SOP either with a cuvette spectrometer or a plate reader (*i.e.* no adaptation of the SOP to other devices). Among the 30 groups who replied to the questionnaire, two groups, originating from Spain and Russia were discarded because they did not agree to perform the OP-AA SOP, and one group from India withdrew from the intercomparison before sample shipment because of a lack of laboratory equipment. While we additionally reached out to collaborators from Chile and Brazil, they did not have sufficient workforce to be able to participate. We acknowledge that geographical repartition of participants is not representative of the entire OP community but we believe that a diversity of environmental/meteorological conditions was encountered by participants.

The Recommendations for future inter-laboratory comparisons was amended with the following sentence lines 725-728: *“Although participants originated from countries with different meteorological and environmental conditions, future ILCs should try to include more diversity in geographical repartition of participating laboratories, to increase the representativity and challenges of the global OP community.”*

L445-446: Uniform area of filter punch should have been used by all the labs because here the purpose was also to check the contributions from filters to OP-AA.

We agree with the reviewer that this would be the ideal case, however this could not be achieved for practical reasons. Since the ILC accepted laboratories using both plate-readers and cuvettes, the latter required large volumes of extracts and therefore large filter surfaces. When performing analyses in triplicate with triplicate extractions, this resulted in providing up to 2.5 cm<sup>2</sup> of a filter only to perform the OP-AA SOP (*i.e.* without providing sample to perform the OP-AA home protocol nor the OP-DTT ILC1 SOP), making it impossible to give the possibility to participants to test other protocols.

Fig. 3: Here, it would be nice to include a comparison of reagent blanks for all labs. Reagent blank means everything is same as samples during the analytical procedure but instead of sample or blank filter extracts, only deionized water was added. From the description given in the methodology section, it is not clear whether 'negative control'

itself is reagent blank or it was just the absorbance correction due to buffer solution. Comparison of reagent blanks and filter blanks would help in assessing the steps introducing variability among different labs. Add this discussion in section 3.4.3.

Thank you for this suggestion. Since samples are extracted in K-PO<sub>4</sub> buffer, the triplicate analysis of K-PO<sub>4</sub> buffer is considered both as a negative control but also as reagent blank, with the average slope in the triplicates subtracted from the slope of the samples (see l. 226-227). Since ultra-pure water is used for AA preparation, a reagent blank consisting in replacing the sample by deionized water would result in pH conditions that are not representative of the samples' analytical conditions, additionally leading to a shift in AA's maximum absorption wavelength (see Sect. 1.4 of the Supplement Material). This is why the reagent blank, called here negative control, is made of K-PO<sub>4</sub> buffer and follows the same procedure as the samples.

We clarified this in the Methods section line 219: "Triplicate analyses of the K-PO<sub>4</sub> buffer were also used as negative controls (or reagent blank) for AA background oxidation in the extraction matrix."

Following the reviewer's suggestion, we have also compared the negative control and sample B (i.e. filter blank) among participants (Figure R1), using the following formula:

$$Deviation = 100 \frac{\overline{OP}_{slope}^{B,i} - \overline{OP}_{slope}^{neg.control}}{\overline{OP}_{slope}^{neg.control}}$$
 where  $\overline{OP}_{slope}^{B,i}$  is the raw (i.e. uncorrected for negative control) average slope of sample B in the 3 replicates of extract i, and  $\overline{OP}_{slope}^{neg.control}$  is the average slope of all replicates of the negative control.

Figure R1 highlights that the magnitude and direction of sample B's deviation from the negative control vary substantially among laboratories. The median (25<sup>th</sup> percentile, 75<sup>th</sup> percentile) of 2 (-26, 26) % suggests that for most participants, the filter blank introduced a marginal effect as compared to the laboratory-dependent conditions.

Figure R1 was added in the Supplement Material, and section 3.4.3 was modified with the addition of the following sentence lines 491-494: "*The deviation of sample B's slope from the average slope of the negative controls (Fig. S10) ranged from -284% to 328%, with a median (25th percentile, 75th percentile) of 2 (-26, 26) %. This suggests that for most participants, the filter blank introduced a marginal effect as compared to the laboratory-dependent conditions.*"

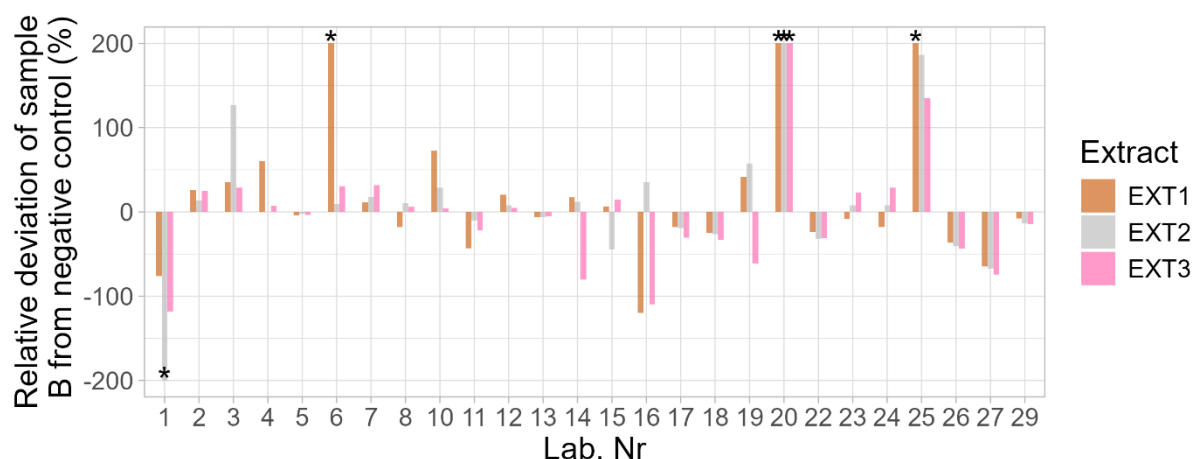


Figure R1. Relative deviation of sample B's slope from the average slope of the negative controls. Black stars represent the discarded replicates to calculate the LOD per laboratory.

L474-475: If the measurements are precise, these factors can cause a systematic deviation and not random deviations.

We agree with the reviewer that for one laboratory and if measurements are precise, all the cited factors, *i.e.* instrument model, stability, calibration accuracy, water quality and protocol compliance should cause systematic deviations. However, in the case of an inter-laboratory comparison, measurements are not necessarily precise and these factors may provoke random deviations when considering the 26 laboratories. In the protocol, we tried to control for within-laboratory systematic errors by providing the reagent powder, by insisting on the importance of calibrating the pipettes (last calibration of the pipettes asked in the Results Excel file), including only two types of spectrophotometers (*i.e.* cuvette based and plate-readers), checking the K-PO<sub>4</sub> pH before launching any analysis. To reduce within-laboratory random errors, the protocol insisted on the importance of washing consumables to avoid uncontrolled effects of contamination of labware, and the organizing laboratory provided readily punched samples to avoid random and systematic errors that could occur during punching. The factors addressed in the paragraph are therefore influenced by both systematic and random errors.

We modified the sentence line 480 that now reads as: “*The investigated analytical variables can be influenced by **both systematic and random deviations**, such as the instrument model or stability, the calibration accuracy, the water quality, or protocol compliance.*”

L475-477: Use of average slope of the negative controls could be another significant source of uncertainty, unless it is very similar on different days. Reagent blank must be run along with every batch of analysis, and respective blanks shall be used for

corrections. It has been observed that this reagent blank vary considerably from day-to-day. If reagent blanks were done everyday, the suggested way can be checked for better comparison among labs.

We thank the reviewer for this comment. Since samples are extracted in K-PO<sub>4</sub> buffer, the triplicate analysis of K-PO<sub>4</sub> buffer is considered both as a negative control but also as reagent blank, with the average slope in the triplicates subtracted from the slope of the samples (see l. 226-227). Since all plate-reader users except one performed the analysis of all samples within one day, the variability of reagent blanks could not be assessed more accurately than with the standard deviation between the three replicates.

Line 219 in the Methods section now read as: “Triplicate analyses of the K-PO<sub>4</sub> buffer were also used as negative controls (**or reagent blank**) for AA background oxidation in the extraction matrix.”

L631-633: Why glass fibre or other types of filters cannot be used? There was no description on the limitation of filter media for OP-AA analysis?

Thank you for these questions. We only advised on the two most commonly used types of filters, quartz fiber filters being the reference material for PM<sub>10</sub> collection in the national air monitoring networks in EU (EN 12341), and Teflon® filters being often use for metal analysis. To the best of our knowledge, no study specifically addressed the effects of glass fiber filters compared to other types of filters in the literature. Salana et al., (2024) compared OP analyzed in filters of different natures: quartz fibers on 1 site, Teflon and glass fibers on 2 other sites, but they did not compare the impact of filter nature in their study. Rapisarda et al. (2015) found reduced cell viability and increased DNA damage and oxidative stress production in human epithelial cells exposed to glass fibers. Since the extraction step of glass fiber filters would potentially destroy the fibers, and since OP aims to reproduce physiological conditions, we do not think that glass fiber filters would be well suited for OP analysis. However, since we do not have formally tested this aspect, we would prefer to avoid mentioning this in the manuscript.

To be consistent with the lack of formal testing of filter nature, we also removed the lines 631-632 of the preprint. The Recommendations for future inter-laboratory comparisons part of the General recommendations was also amended, lines 729-730 with the following sentence: “*Filter nature i.e. quartz fiber, Teflon®, and glass fiber filters could also be compared.*”

L696-700: It would be nice to include reagent blank in each batch of analyses, which shall be suitably subtracted from sample's OP.

We thank the reviewer for this suggestion. We modified the general recommendations section lines 689-690 accordingly, by adding: *“Replicated negative control consisting in replacing sample extract by the K-PO<sub>4</sub> buffer extracting matrix should be run along every batch of analysis.”*

L701-706: For future ILC, I would recommend to use multiple real samples collected from regions dominated by different types of sources e.g., megacities, semi-arid/arid regions, biomass burning dominated regions, pristine environment like Arctic/Antartic or remote mountains, marine or island sites, etc. This will help in assessing the effect of different matrix extractions in different labs. In this section, please add that all lab shall periodically run internal standard (CuSO<sub>4</sub>), and make sure that they are getting an acceptable blank corrected slope (define it). It will help in generation of quality data across the labs, and values can be used/compared for broader scientific purposes.

We thank the reviewer for these suggestions. Although the highest rate of satisfactory results is obtained for CuSO<sub>4</sub>, we state in the manuscript that a coefficient of variation of 40% among all participants is too high to ascribe an OP-AA reference value (see line 630). Instead, we are now suggesting each laboratory to define an acceptable range for an internal standard of their choice, by repeating the analysis several times.

The general recommendations were amended accordingly, by adding in the Calculation step lines 720-723: *“- A chosen internal standard such as CuSO<sub>4</sub> or a dedicated atmospheric sample should be analyzed along each analytical batch. The corrected slope obtained should remain stable over time and fall within a predefined acceptable range established by the laboratory by repeating the measurements of the standard under stable operating conditions.”*

And in the Recommendations for future inter-laboratory comparisons lines 728-729: *“The tested samples could also be chosen to represent this geographical diversity, with samples influenced by more specific sources.”*

## Response to referee n#2:

### Overall assessment:

This manuscript presents a timely and valuable interlaboratory comparison of the ascorbic acid (AA) assay for measuring the oxidative potential (OP) of particulate matter (OP-AA). I found the overall effort impressive, particularly the scale of the exercise involving 26 laboratories. The harmonization of the analytical procedure, including the extraction step, and the detailed assessment of factors contributing to interlaboratory variability are particularly valuable and well justified. The assessment of variability across repeatability, extraction, and between-laboratory components is informative, while the comparison of different analytical approaches provides useful insights for further harmonization of OP-AA measurements.

The study is especially relevant given the continuing challenges in obtaining comparable OP measurements across laboratories and studies. In particular, factors such as PM mass loading and initial AA concentration can substantially influence the reported OP if not appropriately considered. The present effort therefore provides a useful contribution toward improving the consistency and interpretation of OP-AA measurements.

Overall, I consider the study potentially suitable for publication in Atmospheric Measurement Techniques. However, several aspects of the statistical treatment, analytical evaluation, and interpretation would benefit from further clarification before publication. I therefore encourage the authors to address the following points.

### Comments:

- **Lines 270-280 and 395-400: The definition of the proficiency standard deviation and the z-score categories need clarification:**

The manuscript states that the standard deviation for proficiency assessment was fixed at 20% of the assigned value, and the z-score is calculated as:

$$z = (x_i - X) / (0.2X)$$

However, the manuscript subsequently describes  $\pm 20\%$  as the “acceptable range,” while defining  $z$  within  $\pm 2$  as satisfactory.

With the equation used here, a  $\pm 20\%$  deviation corresponds to  $z = \pm 1$ , whereas  $z = \pm 2$  corresponds to  $\pm 40\%$  deviation from the assigned value. Thus, “20% acceptable range” and “satisfactory at  $z$  within  $\pm 2$ ” are not equivalent concepts.

I recommend that the authors clearly distinguish between:

1. the standard deviation used for proficiency assessment (20% of  $X$ ), and
2. the limits used to classify  $z$ -scores as satisfactory, warning or action.

We agree with the reviewer that the use of “acceptable range” instead of “standard deviation used for proficiency assessment” is misleading. We modified the Data evaluation of the Methods section and the General description of the 3.4 section that now read as:

l.279-280: *“This value ~~is represents the maximum acceptable error~~ in accordance with literature estimations of uncertainties for OP-DTT analysis (Molina et al., 2020) and OP-AA variations (Fang et al., 2016).”*

Lines 399-401: *“For samples A, C, D and EF2, the laboratory performances were evaluated as the deviation from the assigned value, ~~with an acceptable range of +/- 20%~~ (Figure 3, assigned values of 0.1210, 0.4674, 0.1047, 78.49 nmol min<sup>-1</sup> µg<sup>-1</sup> for samples A, C, D, EF2, respectively).”*

- **Lines 395-425 and 610-615: The reported assigned values, particularly EF2, need additional contextualization:**

The assigned values reported for samples A, C, D and EF2 are 0.1210, 0.4674, 0.1047 and 78.49 nmol min<sup>-1</sup> µg<sup>-1</sup>, respectively. For EF2, which was introduced specifically as a potential chemical reference, I think it would be useful to provide more information on how the assigned value compares with previously published OP-AA measurements of CuSO<sub>4</sub>·5H<sub>2</sub>O under comparable conditions.

We thank the reviewer for this comment that will provide more insights on sample EF2. Table R1 summarizes the studies reporting AA depletion slopes and we calculated the corresponding OP<sub>m</sub> in nmol min<sup>-1</sup> µg<sup>-1</sup> based on the available information in each study. Values range from 5.17 10<sup>-3</sup> nmol min<sup>-1</sup> µg<sup>-1</sup> to 78.49 nmol min<sup>-1</sup> µg<sup>-1</sup>, the latter being sample EF2's assigned value in this study. A [-2; 2] z-score range corresponds to OP<sub>m</sub><sup>AA</sup> values from 47.09 to 109.89 nmol min<sup>-1</sup> µg<sup>-1</sup>. The study by Expósito et al. (2024) found very low values compared to the others, which they assigned to the presence of chloride in PBS, that was found to decrease AA response in the studies by Pietrogrande et al. (2019, 2022). The studies by Pietrogrande et al. (2019, 2022) applied the protocol published by Visentin et al. (2016) and the reported values in the 3 studies align, with mean ± standard deviation in OP<sub>m</sub><sup>AA</sup> of 33.24 ± 4.55 nmol min<sup>-1</sup> µg<sup>-1</sup>, but they fall outside the acceptable range of the present study. Lower values are consistent with the lower initial AA concentration (100µM). Assuming a pseudo-first order kinetics, the values by Pietrogrande (2022) at 0.17 µM would be of 54.91 nmol min<sup>-1</sup> µg<sup>-1</sup>, falling within the satisfactory z-score zone, but this would not be the case with the value reported by Visentin et al. (2016) in the same conditions. The values published by Sekine et al. (2025) are in closer accordance with the assigned value here, despite having lower initial AA concentration. This is consistent with a reaction in Milli-Q water, avoiding part of copper to be complexed by phosphate. The more acidic pH is also probably shifting the

maximum absorption wavelength and explain some of the differences observed (see Sect. 1.4 of the Supplement material). The present study and the published ones therefore mainly differ in the initial AA concentration. CuSO<sub>4</sub> concentration, and the solution preparation for CuSO<sub>4</sub>, such as phosphate concentration in the reactive medium can also play a role in the differences in AA depletion slopes.

We have added Table R1 as Table S6 in the Supplement Material and we are now discussing the assigned value in the Lessons learned section, lines 623-628: “To the best of our knowledge, sample EF2’s assigned value is the highest reported value for CuSO<sub>4</sub>, with  $OP_m^{AA}$  ranging from  $5.17 \cdot 10^{-3} \text{ nmol min}^{-1} \mu\text{g}^{-1}$  in phosphate buffer saline (Expósito et al., 2024) to  $78.49 \text{ nmol min}^{-1} \mu\text{g}^{-1}$  in the present study. Three similar studies reported an average  $OP_m^{AA}$  of  $33.24 \pm 4.55 \text{ nmol min}^{-1} \mu\text{g}^{-1}$ , which would correspond to a z-score of -2.9. The differences can mainly be attributed to the initial AA concentration (100  $\mu\text{M}$  vs. 133  $\mu\text{M}$ ), since values would be higher when assuming a pseudo-first order kinetics as suggested in previous studies (Expósito et al., 2024; Souza et al., 2025). CuSO<sub>4</sub> concentration and phosphate concentration during the reaction step can also explain these differences (Table S6).”

Table R1. Reported AA depletion slopes and corresponding  $OP_m^{AA}$  for different concentrations of CuSO<sub>4</sub>, 5H<sub>2</sub>O.

| Reference                           | Reagent used                                | [AA] <sub>0</sub> | Concentration in reactive medium          | Reported slope (corrected for reagent blank) | Calculated OP <sub>m</sub> (nmol min <sup>-1</sup> µg <sup>-1</sup> ) | Information/assumptions used for calculations*                                                  |
|-------------------------------------|---------------------------------------------|-------------------|-------------------------------------------|----------------------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Assigned value of EF2 in this study | CuSO <sub>4</sub> , 5H <sub>2</sub> O, ≥99% | 133 µM            | 0.2 µM in K-PO <sub>4</sub> buffer 0.1M   | /                                            | 78.49                                                                 | /                                                                                               |
| (Expósito et al., 2024)             | CuSO <sub>4</sub> , analytical grade        | 200 µM            | 0.34 µM in PBS1X                          | 0.47 µM min <sup>-1</sup>                    | 5.53 10 <sup>-3</sup>                                                 | M(CuSO <sub>4</sub> , 5H <sub>2</sub> O) = 249.68 g mol <sup>-1</sup><br>Final volume = 200 µL  |
|                                     |                                             |                   | 1 µM in PBS1X                             | 1.29 µM min <sup>-1</sup>                    | 5.17 10 <sup>-3</sup>                                                 |                                                                                                 |
| (Pietrogrande et al., 2019)         | CuSO <sub>4</sub> , 5H <sub>2</sub> O, 98%  | 100 µM            | 0.1 µM in Na-PO <sub>4</sub> buffer 0.1M  | 2.33 nmol min <sup>-1</sup>                  | 30.79                                                                 | M(CuSO <sub>4</sub> , 5H <sub>2</sub> O) = 249.68 g mol <sup>-1</sup><br>Final volume = 3.03 mL |
| (Pietrogrande et al., 2022)         |                                             |                   | 0.05 µM in Na-PO <sub>4</sub> buffer 0.1M | 1.20 nmol min <sup>-1</sup>                  | 31.72                                                                 |                                                                                                 |
|                                     |                                             |                   | 0.10 µM in Na-PO <sub>4</sub> buffer 0.1M | 2.28 nmol min <sup>-1</sup>                  | 30.13                                                                 |                                                                                                 |
|                                     |                                             |                   | 0.17 µM in Na-PO <sub>4</sub> buffer 0.1M | 5.31 nmol min <sup>-1</sup>                  | 41.29                                                                 |                                                                                                 |
| (Visentin et al., 2016)             | Cu(II)                                      | 100 µM            | 0.17 µM in Na-PO <sub>4</sub> buffer 0.1M | 3.52 µM min <sup>-1</sup>                    | 32.32                                                                 |                                                                                                 |
| (Sekine et al., 2025)               |                                             | 99 µM             | 0.1 µM in MilliQ water                    | 1.8 µM min <sup>-1</sup>                     | 72.09                                                                 |                                                                                                 |

|  |                                                     |  |                           |                             |       |                                                                                                         |
|--|-----------------------------------------------------|--|---------------------------|-----------------------------|-------|---------------------------------------------------------------------------------------------------------|
|  | CuSO <sub>4</sub> ,<br>5H <sub>2</sub> O,<br>99.5+% |  | 0.2 µM in MilliQ<br>water | 2.9 µM<br>min <sup>-1</sup> | 58.07 | M(CuSO <sub>4</sub> ,<br>5H <sub>2</sub> O = 249.68<br>g mol <sup>-1</sup><br>Final volume =<br>10.1 mL |
|  |                                                     |  | 0.3 µM in MilliQ<br>water | 3.8 µM<br>min <sup>-1</sup> | 50.73 |                                                                                                         |
|  |                                                     |  | 0.5 µM in MilliQ<br>water | 4.8 µM<br>min <sup>-1</sup> | 38.44 |                                                                                                         |

\*This information was used to calculate OP<sub>m</sub><sup>AA</sup> in nmol min<sup>-1</sup> µg<sup>-1</sup> because the unit of the reported data varied.

More importantly, the manuscript reports a CV of 40% among the 26 participants and concludes that this is still too high to establish an OP-AA reference value. Please clarify exactly how this 40% CV was calculated. This clarification is particularly important because EF2 is being considered as a potential future reference material.

The CV of 40% among all participants was calculated by dividing the standard deviation of OP<sub>m</sub> by the average of OP<sub>m</sub> across all participants, using the average of the three replicates obtained by each participant. We clarified this in the corresponding section lines 621-622: “CV of 40% among all 26 participants, using the average of the three replicates for each participant.”

- **Lines 445-455 and 500-520: Signal-to-background and analytical sensitivity.**

The large variability observed for the lower-OP samples deserves further consideration in relation to the signal-to-background conditions of the measurements. Samples A and D show the largest z-score ranges, while their assigned OP values are substantially lower than that of sample C. At the same time, the blank-filter experiment shows substantial differences in estimated LOD among laboratories, particularly between cuvette- and plate-reader users. A larger fraction of cuvette users also reported samples A and C as below their estimated LOD.

I therefore suggest that the authors examine whether the observed laboratory-to-laboratory variability, particularly for the lower-OP samples, is related to signal-to-background ratio or laboratory-specific analytical sensitivity. Since laboratory-wise z-scores and LODs are already available (Figs. 3 and 5), an analysis of the relationship between these parameters could be informative. Reporting the approximate signal-to-background characteristics of the different samples, where possible, would also help clarify whether the larger z-score ranges for the lower-OP samples partly reflect measurements being closer to the analytical detection limit. This could provide a more quantitative basis for interpreting the observed instrument-related differences.

We thank the reviewer for this question. The section 3.4.2 has now been corrected and lines 457-464 reads as:

*“Participants using plate-readers had LODs ( $\text{nmol L}^{-1} \text{min}^{-1}$ ) similar to participants using cuvettes (Figure 5). Therefore, the percentage of samples below the LOD was similar in the two groups, with 19% (i.e. 3 out of 16) vs. 20% (i.e. 2 out of 10) in the plate-reader users vs. cuvette users for sample A, 6% vs. 20% for sample C, 13% vs. 20% for sample D and 0% vs. 10% for sample EF2. Participants in the satisfactory z-score zone did not tend to have lower LOD as compared to participants in the warning or action zones (Table S5) for any of the samples, suggesting that z-scores were not influenced by laboratory-specific sensitivity. This conclusion was not modified when using all of sample B’s replicates for LOD calculation (Fig. S12). The observed similarities suggest that the two methods are equivalent under low signal conditions or low-volume samples.”*

We also tested whether participants falling in the satisfactory z-scores had lower LOD compared to participants in the action and warning zones, for each sample, which was not the case for any of the samples. Table R2 was added in the Supplement Section and we added the following sentence in section 3.4.2, lines 461-463: *“Participants in the satisfactory z-score zone did not tend to have lower LOD as compared to participants in the warning or action zones (Table S5) for any of the samples, suggesting that z-scores were not influenced by laboratory-specific sensitivity.”*

Table R2. Comparison of LOD ( $\text{nmol L}^{-1} \text{min}^{-1}$ ) in participants in the satisfactory vs. warning or action z-score zones for each sample.

|                                                                          | Satisfactory  | Warning or Action | p-value <sup>†</sup> |
|--------------------------------------------------------------------------|---------------|-------------------|----------------------|
| <b>Sample A</b>                                                          |               |                   |                      |
| <b>LOD (<math>\text{nmol L}^{-1} \text{min}^{-1}</math>)<sup>*</sup></b> | 153 (80, 196) | 251 (133, 479)    | 0.2                  |
| <b>N</b>                                                                 | 16            | 10                |                      |
| <b>Sample C</b>                                                          |               |                   |                      |
| <b>LOD (<math>\text{nmol L}^{-1} \text{min}^{-1}</math>)<sup>*</sup></b> | 165 (82, 212) | 145 (118, 402)    | 0.8                  |
| <b>N</b>                                                                 | 18            | 8                 |                      |
| <b>Sample D</b>                                                          |               |                   |                      |
| <b>LOD (<math>\text{nmol L}^{-1} \text{min}^{-1}</math>)<sup>*</sup></b> | 165 (85, 212) | 145 (85, 402)     | >0.9                 |
| <b>N</b>                                                                 | 14            | 12                |                      |

<sup>\*</sup>Wilcoxon rank sum exact test; <sup>†</sup>Median (Q1, Q3);

- Lines 445-455; Figure 5: The LOD calculation deserves further methodological clarification:**

The theoretical LOD is calculated as  $3\sigma_B$ , where  $\sigma_B$  is obtained from nine blank-filter measurements, after allowing up to three outliers to be removed while retaining at least six results. I understand the rationale for using the blank filter as a procedural and instrumental noise. However, removing up to one-third of the observations before estimating the standard deviation could substantially reduce the estimated  $\sigma_B$ , and therefore produce an optimistic LOD. The authors should clarify the exact criterion used to identify the up-to-three outliers, and whether the LODs change substantially if all nine

measurements are retained. This is important because the LOD is subsequently used to interpret differences between cuvette and plate-reader performance.

We agree with the reviewer that LOD calculation could be clarified. Several approaches were tested for outlier identifications at the participants' level:

- $OP_{slope,ri} \geq p75 + 1.5 (p75 - p25) \mid OP_{slope,ri} \leq p25 - 1.5 (p75 - p25)$  with  $OP_{slope,ri}$  the  $OP_{slope}$  of  $i^{th}$  replicate of the negative control,  $p75$  and  $p25$  being the 75<sup>th</sup> and 25<sup>th</sup> percentiles in the replicates of the negative control.
- $OP_{slope,ri} \geq average(OP_{slope}) + sd(OP_{slope}) \mid OP_{slope,ri} \leq average(OP_{slope}) - sd(OP_{slope})$  with  $OP_{slope,ri}$  the  $OP_{slope}$  of  $i^{th}$  replicate of the negative control,  $average(OP_{slope})$  and  $sd(OP_{slope})$  the average and standard deviation of all the replicates of the negative control.
- $OP_{slope,ri} \geq p(99) \mid OP_{slope,ri} \leq p(1)$  with  $OP_{slope,ri}$  the  $OP_{slope}$  of  $i^{th}$  replicate of the negative control,  $p99$  and  $p1$  being the 99<sup>th</sup> and 1<sup>st</sup> percentiles in the replicates of the negative control.

However, all these approaches were also suppressing repeatable values, leading to a more pessimistic LOD. We therefore opted for a customized approach, using Figure R2 to identify values that could be affected by technical problems during the analytical procedure. Figure R2 shows the slopes obtained in each replicate of each extract (in nmol min<sup>-1</sup>), showing that some replicates strongly deviated for some participants. Several participants reported difficulties to pipet one replicate per extract because of the low volume of extract left or because of the turbidity directly following the end of extraction. Therefore, the removal of up to three values enables to take this feedback into account for LOD calculation, although we acknowledge that it may produce an optimistic LOD. Figure R2 was added to the Supplement Material as Figure S11, and it is now referenced in the corresponding paragraph line 456: “A maximum of 3 outliers were removed from each set of results, and at least the 6 most reproducible results were kept (Fig. S11). [...]”

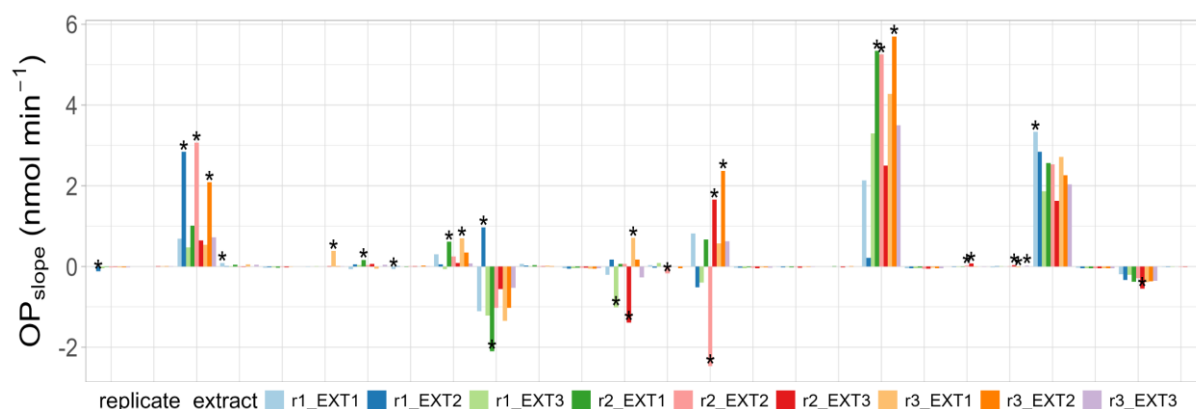


Figure R2. Sample B's  $OP_{slope}$  for each extract and each replicate. Black stars represent the discarded replicates to calculate the LOD per laboratory.

Section 3.4.2 has now been modified, calculating the LOD in  $nmol L^{-1} min^{-1}$  instead of  $nmol min^{-1}$  that was introducing a bias between cuvettes and plate-readers. Figure R3 compares the LOD in  $nmol L^{-1} min^{-1}$  obtained after discarding replicates (panel A) with the LOD calculated with all replicates of all extracts (panel B) and shows that although the LOD was improved for a few participants, the difference between cuvettes and plate-readers was not modified by the approach. Moreover, Table R3 compares the percentages of samples below the LOD depending on the LOD calculation and shows that this percentage was modified for 7 laboratories, among which 4 are using plate-readers and the other 3 are using cuvettes. Therefore, the customized approach did not significantly improve the results of participants using one type of device compared to the others. To avoid any confusion for the readers, we did not add Table R3 to the manuscript.

Figure R3 panel B was added as Figure S12 in the Supplement, the overall discussion of LODs lines 456-465 has been modified to:

*“Participants using plate-readers had LODs ( $nmol L^{-1} min^{-1}$ ) similar to participants using cuvettes (Figure 5). Therefore, the percentage of samples below the LOD was similar in the two groups, with 19% (i.e. 3 out of 16) vs. 20% (i.e. 2 out of 10) in the plate-reader users vs. cuvette users for sample A, 6% vs. 20% for sample C, 13% vs. 20% for sample D and 0% vs. 10% for sample EF2. Participants in the satisfactory z-score zone did not tend to have lower LOD as compared to participants in the warning or action zones (Table S5) for any of the samples, suggesting that z-scores were not influenced by laboratory-specific sensitivity. This conclusion was not modified when using all of sample B's replicates for LOD calculation (Fig. S12). The observed similarities suggest that the two methods are equivalent under low signal conditions or low-volume samples.”*

Table R3. Comparison of percentage of replicates below LOD that differ when using LOD calculation with all of sample B's replicates compared with LOD calculation after removing up to one third of replicates.

| Lab. Nr | Device | Sample | LOD <sub>all</sub><br>( $nmol L^{-1} min^{-1}$ ) | Percentage of samples below LOD <sub>all</sub> | LOD <sub>cus</sub><br>( $nmol L^{-1} min^{-1}$ ) | Percentage of samples below LOD <sub>cus</sub> |
|---------|--------|--------|--------------------------------------------------|------------------------------------------------|--------------------------------------------------|------------------------------------------------|
| 1       | PR     | A      | 562                                              | 100%                                           | 154                                              | 11%                                            |
|         |        | D      |                                                  | 100%                                           |                                                  | 0%                                             |

|    |         |     |      |      |      |      |
|----|---------|-----|------|------|------|------|
| 3  | Cuvette | A   | 1151 | 89%  | 212  | 0%   |
|    |         | D   |      | 22%  |      | 0%   |
| 4  | PR      | A   | 595  | 100% | 456  | 44%  |
| 6  | PR      | A   | 2131 | 22%  | 136  | 0%   |
|    |         | D   |      | 100% |      | 0%   |
| 15 | PR      | A   | 1160 | 78%  | 624  | 0%   |
|    |         | D   |      | 100% |      | 0%   |
| 16 | Cuvette | A   | 1545 | 67%  | 654  | 56%  |
|    |         | C   |      | 22%  |      | 22%  |
|    |         | D   |      | 56%  |      | 56%  |
|    |         | EF2 |      | 11%  |      | 11%  |
| 20 | Cuvette | A   | 1986 | 100% | 1569 | 100% |
|    |         | C   |      | 100% |      | 78%  |
|    |         | D   |      | 100% |      | 78%  |

Note:  $LOD_{all}$ : calculated using all replicates;  $LOD_{cus}$ : calculated after removing up to 1/3 of replicates

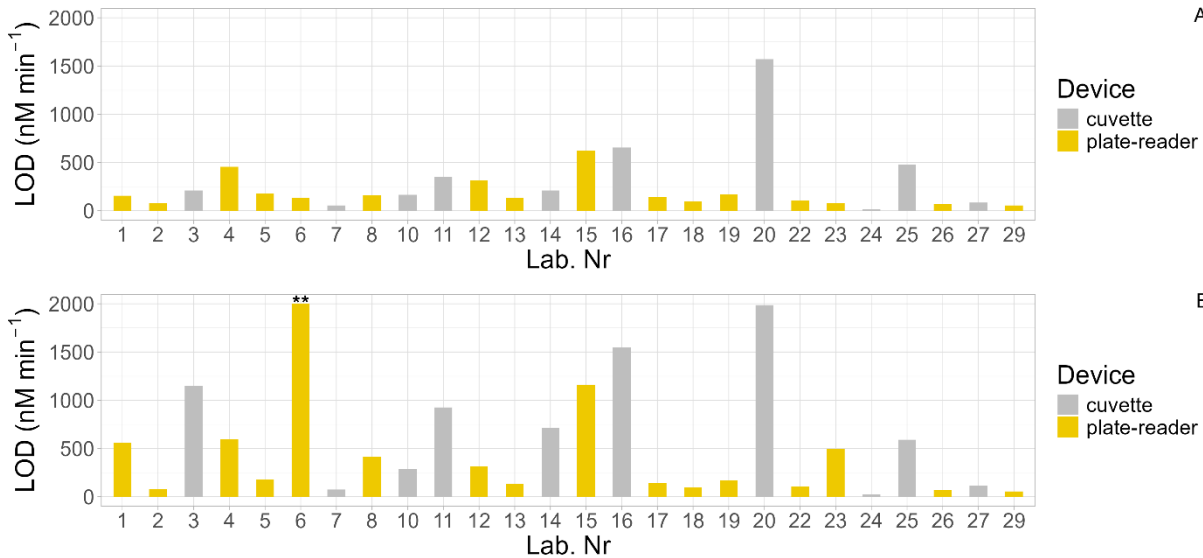


Figure R3. Comparison of the LODs ( $\text{nmol L}^{-1} \text{min}^{-1}$ ) calculated after removing up to 1/3 of the replicates (A) and LODs calculated with all replicates of all extracts (B). Results were cropped at  $LOD \leq 2000 \text{ nmol L}^{-1} \text{min}^{-1}$ , and stars represent bars that exceed this value.

- **Lines 475 and throughout Sections 3.4.3 and 3.5: The calibration performance should be examined more explicitly:**

The manuscript identifies calibration accuracy as a possible source of variability but does not appear to provide sufficient quantitative information on calibration performance among laboratories. Since the calibration curve is directly used to calculate OP, calibration quality could be a major contributor to between-laboratory differences. I suggest providing, if available, information on the calibration slopes, intercepts, and linearity. Comparison of calibration quality within each instrument configuration, together with its relationship to laboratory z-scores, could be very informative.

We thank the reviewer for this comment. The slope of the calibration curve is not a sufficient criterion to assess whether a participant performed well, only a standard could be used to assess whether the concentration calculated with the calibration curve is precise. Likewise, a calibration can be “perfectly wrong”. In fact,  $R^2$  of the calibration curves Absorbance = f(AA concentration) was  $>0.99$  for 22 out of 26 participants, and except Lab. 16 who reported a problem with the provided AA reagent and obtained an  $R^2$  of 0.13, the minimum  $R^2$  was 0.93. This shows that the  $R^2$  cannot predict the performance of participants. Since there is no AA standard that could be analyzed by each participant to assess their respective calibration accuracy, we opted for the comparison of the theoretical vs. calculated AA concentration at time zero as an alternative approach.

The information on calibration slopes and intercepts will be provided in the Supplemental Excel file that also compiles participants’ results (not made available at the initial submission).

- **Lines 485-520: The relationship between  $[AA]_0$  and laboratory performance is potentially one of the strongest findings, but the interpretation could be strengthened:**

The manuscript reports differences in performance between laboratories whose measured  $[AA]_0$  was within versus outside the theoretical range of 120-146  $\mu\text{M}$ . Please clarify the basis for the proposed 120-146  $\mu\text{M}$  range (10%) and whether this range can be recommended as a practical QC criterion for identifying potential analytical problems.

Thank you for this question. The 10% around the theoretical  $[AA]_0$  was proposed as an operational range to assess deviations from  $[AA]_0$  and was found to be a strong indicator for laboratory performances. However, three participants are outside the range but still obtained satisfactory results for some samples, showing that small deviations may still lead to good results. Therefore, we do not think that this range should be recommended as a practical QC criterion at this stage.

The General recommendations already include this aspect in the Calculation step paragraph, lines 716-719: *“Before running a long series of analyses, the ILC2 demonstrated the importance of performing preliminary controls: ensuring the linearity of the calibration curve, checking for coherence between  $[AA]_{th}$  and  $[AA]_0$ , and verifying the consistency and low values of the slope of the negative controls and/or blank filters. Validating these preliminary steps should limit the risk of incoherent results.”*

- **Section 3.5, Lines 525-545: clarify the terminology and comparison of the two OP calculation approaches:**

This section is scientifically interesting, but the terminology used for the two calculation approaches is somewhat difficult to follow; e.g., AA calibration curve, mathematical determination of OP, and concentration-based methods (L525-528).

I suggest defining the two approaches explicitly at the beginning of the section and using consistent terminology throughout, for example:

1. calibration-curve-based OP calculation, and
2. concentration-based OP calculation using measured [AA]<sub>0</sub>

We thank the reviewer for this comment. Lines 525-528 aim to explicitly describe the two calculation approaches compared by Souza et al. (2025). We think that together with the details provided in the Methods section, these lines cannot be further clarified.

- **Lines 630-635: The recommendation concerning methanol pre-wetting of Teflon filters needs stronger support:**

The manuscript recommends using a drop of methanol to wet Teflon filters before extraction because of their hydrophobic properties. I think this recommendation requires some additional justification, as the methanol pre-wetting step could influence the recovery of certain hydrophobic or redox-active constituents during a subsequent aqueous extraction. This is distinct from performing an extraction using methanol, but the authors should nevertheless provide evidence or literature support demonstrating that the proposed pre-wetting step does not introduce a systematic change in measured OP-AA. If the recommendation is retained, I suggest specifying approximately how much methanol is intended and the time between pre-wetting and addition of the aqueous extraction solution.

We agree with the reviewer that this paragraph lacks of justification. Since this aspect was not formally tested in the framework of the inter-laboratory comparison, we removed the sentence from the manuscript, and amended the Recommendations for future inter-laboratory comparisons part of the General recommendations lines 729-730 with the following sentence: *“Filter nature i.e. quartz fiber, Teflon®, but also glass fiber filters could also be compared.”*

- **Section 2.3, lines 230–240:**

I would appreciate some clarification regarding the preparation and distribution of the filter punches. Samples A-D were collected on individual ambient filters, while the ILC involved 26 laboratories and multiple punches and extractions per sample. The manuscript states that sufficient punches were provided according to the participants’ declared number of extractions, but it would be useful to explain how the required

number of punches were prepared from each original filter and how these punches were distributed among the participating laboratories. This is relevant for understanding how consistency of the test material was maintained across all participants.

Thank you for this comment. Each individual ambient filter was punched with a 0.5 cm diameter tempered steel punch. The number of punches per laboratory and per sample was calculated depending on the declared instrument and protocol to be applied. Therefore, each laboratory received a package with 4 Petri dishes containing 3 to 36 punches and protected from light with aluminum foil.

The homogeneity, stability and conservation tests checked by the Joint Research Centre aimed to assess the consistency of tested material among participants. The homogeneity assessed OP variations for punches randomly punched on the filter and extracted several times each. The conservation test also mimicked the package preparation, with the required number of punches placed into Petri dishes covered with aluminum foil and placed in the fridge until analysis. Results of both tests showed very good homogeneity (<4%) and coefficients of variation ranging from 3.6 to 6.5% showing that the tested material should be consistent across all participants (see section 3.2).

We have added additional information in the Methods section lines 231-236 that now reads as: *“For OP of PM, punches (**tempered steel**, 0.5 cm in diameter) of impacted or blank quartz fiber filters were provided to the participants in labeled Petri dishes protected from light with aluminum foil (4 samples, named A, B, C, D). Based on participants’ declaration regarding their instrument (i.e. plate-reader or cuvette-based spectrometer) and the protocols they intended to follow (i.e. whether they were to apply their “home” protocols), the organizer provided enough filter punches to be able to perform the declared number of extractions (**final number of punches in each Petri dish ranging from 3 to 36**).”*

## Bibliography

Calas, A., Uzu, G., Martins, J. M. F., Voisin, D., Spadini, L., Lacroix, T., and Jaffrezo, J.-L.: The importance of simulated lung fluid (SLF) extractions for a more relevant evaluation of the oxidative potential of particulate matter, *Scientific Reports*, 7, 11617, <https://doi.org/10.1038/s41598-017-11979-3>, 2017.

Expósito, A., Markiv, B., Santibáñez, M., Fadel, M., Ledoux, F., Courcot, D., and Fernández-Olmo, I.: Ascorbate oxidation driven by PM<sub>2.5</sub>-bound metal(loid)s extracted in an acidic simulated lung fluid in relation to their bioaccessibility, *Air Qual Atmos Health*, 17, 177–189, <https://doi.org/10.1007/s11869-023-01436-8>, 2024.

Fang, T., Verma, V., Bates, J. T., Abrams, J., Klein, M., Strickland, M. J., Sarnat, S. E., Chang, H. H., Mulholland, J. A., Tolbert, P. E., Russell, A. G., and Weber, R. J.: Oxidative potential of ambient water-soluble PM<sub>2.5</sub> in the southeastern United States: contrasts in sources and health associations between ascorbic acid (AA) and dithiothreitol (DTT) assays, *Atmospheric Chemistry and Physics*, 16, 3865–3879, <https://doi.org/10.5194/acp-16-3865-2016>, 2016.

Liu, Q. and Schauer, J. J.: The oxidative potential of atmospheric particulate matter as an indicator of health risk of air pollution: a review, *Environ Chem Lett*, 24, 229–250, <https://doi.org/10.1007/s10311-025-01882-z>, 2026.

Molina, C., Andrade, C., Manzano, C. A., Richard Toro, A., Verma, V., and Leiva-Guzmán, M. A.: Dithiothreitol-based oxidative potential for airborne particulate matter: an estimation of the associated uncertainty, *Environ Sci Pollut Res*, 27, 29672–29680, <https://doi.org/10.1007/s11356-020-09508-3>, 2020.

Pietrogrande, M. C., Bertoli, I., Manarini, F., and Russo, M.: Ascorbate assay as a measure of oxidative potential for ambient particles: Evidence for the importance of cell-free surrogate lung fluid composition, *Atmospheric Environment*, 211, 103–112, <https://doi.org/10.1016/j.atmosenv.2019.05.012>, 2019.

Pietrogrande, M. C., Romanato, L., and Russo, M.: Synergistic and Antagonistic Effects of Aerosol Components on Its Oxidative Potential as Predictor of Particle Toxicity, *Toxics*, 10, 196, <https://doi.org/10.3390/toxics10040196>, 2022.

Rapisarda, V., Loreto, C., Ledda, C., Musumeci, G., Bracci, M., Santarelli, L., Renis, M., Ferrante, M., and Cardile, V.: Cytotoxicity, oxidative stress and genotoxicity induced by glass fibers on human alveolar epithelial cell line A549, *Toxicology in Vitro*, 29, 551–557, <https://doi.org/10.1016/j.tiv.2014.12.010>, 2015.

Salana, S., Yu, H., Dai, Z., Subramanian, P. S. G., Puthussery, J. V., Wang, Y., Singh, A., Pope, F. D., Leiva G., M. A., Rastogi, N., Tripathi, S. N., Weber, R. J., and Verma, V.: Inter-continental variability in the relationship of oxidative potential and cytotoxicity with PM<sub>2.5</sub> mass, *Nat Commun*, 15, 5263, <https://doi.org/10.1038/s41467-024-49649-4>, 2024.

Sekine, H., Ito, H., and Sekine, Y.: Effect of Werner-Type Complex Formation of  $\text{Cu}^{2+}$  and  $\text{Fe}^{2+}$  on Oxidative Potentials Assessed Using Ascorbic Acid Assay, *Atmosphere*, 16, 192, <https://doi.org/10.3390/atmos16020192>, 2025.

Souza, E. J. dos S., Uzu, G., Fomba, K. W., Dominutti, P. A., Mhadhbi, T., Jaffrezo, J.-L., and Herrmann, H.: A comparative study of methods for calculating the oxidative potential (OP) of atmospheric particulate matter, *Environ. Sci.: Atmos.*, 5, 889–905, <https://doi.org/10.1039/D5EA00025D>, 2025.

Visentin, M., Pagnoni, A., Sarti, E., and Pietrogrande, M. C.: Urban  $\text{PM}_{2.5}$  oxidative potential: Importance of chemical species and comparison of two spectrophotometric cell-free assays, *Environmental Pollution*, 219, 72–79, <https://doi.org/10.1016/j.envpol.2016.09.047>, 2016.