We are grateful for the reviewer's critical review and suggestions. In the following, we addressed the specific comments one by one. The corresponding changes in the text are highlighted in a marked-up version of the manuscript accordingly. Hereinafter, we structure our responses to the comments in the following sequence:

Referee's comments/questions in bold font.
Author's response in regular font.
Proposed changes in italic font.

Main comments

General
I think, this paper and your (and your colleagues') future work would benefit from stating, which of the characterizations need to be performed regularly, before every experiment and which data can be re-used from your publication as-is. I suggest to describe a "best-practice procedure" in this paper, that you can refer to in the future.

We agree that such a paragraph would be useful and added it to the manuscript as section 4.1.4 (within the chamber characterization chapter in the results/discussion).

Added text:

4.1.4 Best Practice for Future Work

For future work, we propose a best practice procedure to re-evaluate chamber properties when experimental conditions change and time progresses.

- *The qualitative pattern of the solar simulator emission spectrum is stable over time, yet we found a difference in the absolute photon flux by about a factor of 3 over the course of roughly one decade (Fig. 4a). Therefore, the spectrum should be recorded regularly. More frequent checks are needed when there is high experimental throughput or when new HMI lamps are inserted, since each lamp is unique and metal depositions on the glass lead to a relatively quick loss of intensity at the beginning of the lamp lifetime (Bleicher, 2012).*
- *Wall losses need to be re-assessed (i) for experiments operated at different temperatures, (ii) for experiments performed in humid air, (iii) for new VOC photooxidation systems with different compound structures and properties, (iv) for new chambers when the Teflon body was replaced (see below), and (v) for longer periods of time and more complex wall interactions if the experimental duration is increased.*
- *Prior to every single experiment, the chamber needs to be carefully cleaned and evaluated for possible artefacts and contamination by recording chamber blank values. For each set of experiments, at least one control experiment without the VOC precursor should be performed. As the chamber is completely evacuated at the end of any experiment, it should still be empty when the next experiment is started. If this is not the case, leaks in the chamber body need to be fixed. A new Teflon body should be constructed after working with high concentrations of sticky compounds and/or NO_x to avoid carry-over or the unintended formation of photochemically relevant species such as HONO (Bell et al., 2023; Rohrer et al., 2005).*

The following paper was added to the bibliography:

Bell, D., Doussin, J.-F., and Hohaus, T.: Preparation of Simulation Chambers for Experiments, in: A Practical Guide to Atmospheric Simulation Chambers, edited by: Doussin, J.-F., Fuchs, H., Kiendler-Scharr, A., Seakins, P., and Wenger, J., Springer International Publishing, Cham, 113–127, https://doi.org/10.1007/978-3-031-22277-1_3, 2023.

UV part of the bandpass-filtered solar simulator spectrum (section 4.1.1)
As you measure ozone, you could use the photolysis of ozone alone in the chamber to double-check the scaling of the old UV-spectrum. Currently, the scaling of the old dataset is motivated by the comparison with the new dataset at higher wavelength, but it is not clearly tested, if the simple scaling of the old data is correct.

With this comment, you highlight a technical issue that we came across because the new spectroradiometer does not completely overlap with the range of wavelength of the DOAS, which was used previously, leaving a gap in the lower UV-range (λ = 262 – 325 nm). Instead of only reporting the new spectrum, we decided to make use of the previously recorded spectrum, as the qualitative features were very similar in both the old and new spectrum. To check the validity of the proposed scaling of the old dataset for the lower UV-range, we performed two actinometric experiments with NO_2 and methylglyoxal, respectively.

Here, you propose to use the photolysis of ozone as an additional check. This requires quenching of the formed $O(^1D)$ to suppress the formation of $O(^3P)$ and the ultimate re-formation of O_3 via the reaction of $O(^3P)$ with O_2 (e.g. Müller et al., 1995, https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/95GL00203). Due to the risk for carry-over and contamination, we prefer not to add large quantities of $O(^1D)$ reagents such as N_2O into our chamber. However, in our N_2/O_2 matrix, the O_3 photolysis rate is very difficult to evaluate due to the cycling of O_3, O_2, $O(^1D)$, and $O(^3P)$. With all these reactions, the photolysis rate of ozone would need to be calculated by a chemical box model, and would be associated with a high uncertainty due to the numerous assumptions that would play into the final result, especially so as the main O_3 photolysis channel switches from direct $O(^1D)$ production to direct $O(^3P)$ production within the spectral range of the scaled spectrum.

We therefore believe that our NO_2 and methylglyoxal actinometric experiments serve the purpose of validating the (scaled) spectrum sufficiently. We realize that our text might have not been clear enough on the meaning of these actinometric experiments. We therefore made the following changes:

Firstly, we adapted the method part.

Old version (section 3.1.1, last paragraph):
To validate the recorded spectrum, we performed two actinometric experiments.

New version (section 3.1.1, last paragraph):
To validate the recorded spectrum as well as the scaled spectrum in the UV range, we performed two actinometric experiments.

In the Results/Discussion section, we already included a comparison of the calculated and empirical NO_2 and methylglyoxal photolysis rate. We now highlighted the qualitative resemblance of the old and the new spectrum, and also added a statement to contextualise this comparison and to make an assessment of the validity of the spectrum (including the scaled spectrum).

Old version (Sect. 4.1.1, 1st paragraph):
This difference could relate to the specific emission and age of each lamp, or to the vertical distance at which the previous spectrum was recorded. Using the newly measured spectrum (λ = 325 – 1000 nm) and the spectrum scaled from the old data set (λ = 262 – 325 nm), the NO_2 photolysis rates calculated theoretically ($2.07 \pm 0.79 \times 10^{-2}$ s^{-1}) and derived experimentally ($2.32 \pm 1.99 \times 10^{-2}$ s^{-1}, Fig. S2) agree within 12 %. For methylglyoxal, the experimental photolysis rate ($3.66 \pm 0.56 \times 10^{-4}$ s^{-1}, Fig. S3) is 21 % higher than the calculated value ($3.02 \pm 1.15 \times 10^{-4}$ s^{-1}), yet the uncertainties of these rates overlap.

New version (Sect. 4.1.1, 1st paragraph):
This difference could relate to the specific emission and age of each lamp, or to the vertical distance at which the previous spectrum was recorded. Meanwhile, the same qualitative features are visible in both spectra. Using the newly measured spectrum (λ = 325 – 1000 nm) and the spectrum scaled from the old data set (λ = 262 – 325 nm), the NO_2 photolysis rates calculated theoretically ($2.07 \pm 0.79 \times 10^{-2}$ s^{-1}) and derived experimentally ($2.32 \pm 1.99 \times 10^{-2}$ s^{-1}, Fig. S2) agree within 12 %. For methylglyoxal, the experimental photolysis rate ($3.66 \pm 0.56 \times 10^{-4}$ s^{-1}, Fig. S3) is 21 % higher than the calculated value ($3.02 \pm 1.15 \times 10^{-4}$ s^{-1}), yet the uncertainties of these rates overlap. Hence, the updated spectrum with the scaled data in the UV range allows us to reliably derive photolysis rates in the BATCH Teflon chamber.

With this change, we feel that the structure of the following paragraphs within section 4.1.1 would be more intuitive if the photolysis rates of the analytes are mentioned next. Therefore, we changed the order of the paragraphs as follows:

Old version (sect. 4.1.1 structure):

1) *"The measured emission spectrum ..."*
2) *"By design, ..."*
3) *"The emission of the solar simulator ..."*
4) *"The calculated photolysis rates ..."*

New version (sect. 4.1.1 structure):

1) *"The measured emission spectrum ..."*
2) *"The calculated photolysis rates ..."*
3) *"The emission of the solar simulator ..."*
4) *"By design, ..."*

Wall losses (section 4.1.2):
Since your chamber can be temperature-controlled over a large temperature-range, I suggest to test wall losses at different temperatures (shift of vapor pressure – this would be nice to set into relation with existing parametrizations), but also at different humidities (effect of hygroscopicity can play a role here). Also, since your measured wall losses seem to not level off, I would also suggest to test the maximal wall loss rate with sulfuric acid.

We agree that temperature, humidity, and time/equilibration all play an important role for wall loss rates in chamber experiments in general terms. For this study, however, these effects have limited relevance: (i) to determine formation yields from loss-corrected concentrations, we only evaluated the experiments at 298 K, which is also the temperature for which we have determined the wall loss rates, (ii) all experiments were performed in dry air, (iii) all experiments were short enough to describe losses to the wall as a continuous process.

We realize that the temperature, humidity, and experimental duration might change in future applications. However, we feel that it is beyond the scope of this paper to address the effect of all these parameters on the wall loss rates, and that this would be a paper on its own. Still, we acknowledge that wall losses need to be re-defined for other temperatures or humidity levels if such experiments are to be performed/evaluated in the future. We added a statement addressing this need in the "best practice" paragraph which we included in response to the first comment. Also, we specifically pointed out that the wall losses that we describe in this work are valid only for the conditions as described above (dry, T = 298 K, short duration).

Old version methods (3.1.2):
To obtain the individual wall losses, a solution containing all authentic standards was injected into the chamber in a N_2 matrix (c = 12.3×10^{10} molecules cm^{-3}, see SPME-GC-MS calibration in Sect. 3.2.1).

New version methods (3.1.2):
To obtain the individual wall losses, a solution containing all authentic standards was injected into the chamber in a dry N_2 matrix at T = 298±1 K (c = 12.3×10^{10} molecules cm^{-3}, see SPME-GC-MS calibration in Sect. 3.2.1).

Old version results/discussion (4.1.2):
The wall loss rates which were determined empirically in this work (Table S8) ranged between $4.54\pm1.80\times10^{-6}$ s^{-1} (benzaldehyde) and $8.53\pm0.68\times10^{-5}$ s^{-1} (p-hydroxybenzaldehyde).

New version results/discussion (4.1.2):
The wall loss rates which were determined empirically in this work for dry air at T = 298±1 K (Table S8) ranged between $4.54\pm1.80\times10^{-6}$ s^{-1} (benzaldehyde) and $8.53\pm0.68\times10^{-5}$ s^{-1} (p-hydroxybenzaldehyde).

Added text results/discussion (end of section 4.1.2):
It should be noted that the presented wall loss rates as well as their parameterization are valid only for dry conditions and a temperature of 298 K, but may differ if experimental conditions change (Grosjean, 1985; Zhang et al., 2015). Furthermore, both the time frame over which we determined the wall losses empirically and the conducted photooxidation experiments were limited to three to four hours. On longer time scales, desorption processes, wall saturation, and equilibrium between the gas phase and the chamber walls need to be taken into account more specifically (Yeh and Ziemann, 2015; Zhang et al., 2015).

Added text within new best practice part (section 4.1.4):
Wall losses need to be re-assessed (i) for experiments operated at different temperatures, (ii) for experiments performed in humid air, (iii) for new VOC photooxidation systems with different compound structures and properties, (iv) for new chambers when the Teflon body was replaced (see below), and (v) for longer periods of time and more complex wall interactions if the experimental duration is increased.

Minor comments

ll. 120- ... : Please give a temperature range for which given toluene + OH reaction rate and abstraction/addition channel ratio are valid

We added the temperature for the initial reaction of toluene with OH radicals.

Old version (Sect. 2, 1st paragraph):
Toluene reacts with OH radicals at a rate of about $k = 5.6\times10^{-12}$ molecules^{-1} cm^{3} s^{-1} (IUPAC, 2024) either via addition of the OH radical to the aromatic ring structure or via H abstraction from the substituted methyl group.

New version (Sect. 2, 1st paragraph):
Toluene reacts with OH radicals at a rate of about $k = 5.6\times10^{-12}$ molecules^{-1} cm^{3} s^{-1} at $T = 298$ K (IUPAC, 2024), either via addition of the OH radical to the aromatic ring structure or via H abstraction from the substituted methyl group.

We agree that the different Arrhenius behaviours of the addition and abstraction channel will result in a dependence of the channel ratios on temperature, and that this should not be simplified without further explanation. As we do not focus on temperature variations in this paper and the text is already quite long, we prefer to not go into too much detail here, and would instead simply delete this sentence, and focus on the MCM representation which likewise implies a dominance of the addition channel.

Old version (Sect. 2, 1st paragraph):
The addition pathway is dominant with a branching ratio in the range of 0.85 – 0.93 (Atkinson et al., 1980; Hu et al., 2007; Wu et al., 2014). The MCMv3.1 (Bloss et al., 2005) distinguishes between four channels in the primary chemistry of toluene (Fig. 1).

New version (Sect. 2, 1st paragraph):
The MCMv3.1 (Bloss et al., 2005) distinguishes between four channels in the primary chemistry of toluene (Fig. 1).

Table S1: table S1 also needs the name of the species (at least in the table description), not just the short form, as these short forms were not introduced yet. Not in the main text until this point, nor in the SI.

We agree that this information is missing in the SI up to this point. As the space in the table is already limited, we instead added a note explaining the full name of the species and their abbreviations at the very beginning of the SI.

Added text (SI, between TOC and Section S1):
Hereinafter, the following abbreviations are used for toluene photooxidation products and their internal standards (here sorted according to their retention time, as in Table 1 in the main text): phenol-d_6 (PHE-d_6), o-cresol (OCR), m-cresol (MCR), p-cresol (PCR), benzyl alcohol (BOH), o-nitrotoluene (ONT), (nitromethyl)benzene (NMB), m-nitrotoluene (MNT), benzoic acid (BAC), p-methylcatechol (PMC), glycolaldehyde (GAL), nitrocresols (NCR), pyruvic acid (PAC), acetophenone-d_8 (APH-d_8), benzaldehyde (BAL), glyoxal (GLY), methylglyoxal (MGL), p-hydroxybenzaldehyde (PHB).

l. 230: here I wondered, how long is one experiment roughly? Information on this comes much later in the manuscript, but would be interesting already at a much earlier stage, e.g. here.

We agree that the information about the total experimental duration (incl. both irradiated and dark periods) comes too late in the manuscript and is only indirectly mentioned (such as in Fig. 3). We therefore included a statement about the total experimental duration in the caption of Table 2 in the method section where the experiments are first described.

Old version (caption Table 2):
The timing of the first (and all consecutive) SPME-GC-MS samples was varied to better constrain formation yields.

New version (caption Table 2):
The timing of the first (and all consecutive) SPME-GC-MS samples was varied to better constrain formation yields. For all experiments, we obtained 4 SPME-GC-MS samples, resulting in total experimental durations of about three hours.

l. 265: “SIM” has not been explained yet

The abbreviation “SIM” has been explained a bit earlier on the previous page (“The GC-MS was operated with splitless injections, a standard HP-5MS column ramping from a temperature of 45 °C to 280 °C, and in both scan and selected ion monitoring (SIM) mode.”) in line 230 of the original manuscript version. Therefore, we would prefer not to repeat the explanation of the abbreviation.

[no change made]

ll. 339/340: does that mean, the PTR-MS was calibrated only once, due to the complication of focusing on the oxygenated species? While it makes a lot of sense to calibrate these species individually, I suggest to use gas-standards with e.g. a set of hydrocarbons, ketones, siloxanes (...) to allow for more frequent calibrations to monitor variations in its transmission curve.

For toluene, the PTR-ToF-MS was calibrated using the mean signal of the start concentration across all experiments. We also have a gas-standard for PTR-MS calibrations, but we preferred to use the chamber calibration as it better resembles the experimental conditions. We did not observe significant variations in the response over the experimental time frame. We tried to formulate this procedure more explicitly and hope that this change will enhance clarity.

Old version (3.2.2, third paragraph):
For toluene, we evaluated the measured signals in the filled chamber after equilibration and prior to the ignition of the solar simulator across all experiments (Sect. 3.3). These experiments were carried out over a period of 6 weeks, during which time the sensitivity of the instrument is not susceptible to substantial drift. We calibrated toluene using the mean signal for its calculated start concentration of $c = 3.79\times10^{12}$ *molecules* cm^{-3} *and the blank value of the cleaned chamber. The overall quantification error was 10 %, calculated as propagation of the instrumental error (0.30 %) and the experimental error (10 %, variability of monitored start concentrations).*

New version (3.2.2, third paragraph):
For the toluene calibration, we evaluated the measured signal in the filled chamber after equilibration and prior to the ignition of the solar simulator. As we always used the same start concentration of toluene, we averaged the measured response across all experiments (Sect. 3.3). These experiments were carried out over a period of 6 weeks. During this time, we did not observe a substantial drift in the sensitivity of the instrument. We performed a two-point calibration of toluene using the mean signal for its calculated start concentration of $c = 3.79\times10^{12}$ *molecules* cm^{-3} *and the blank value of the cleaned chamber. The overall quantification error was 10 %, calculated as propagation of the instrumental error (0.30 %) and the experimental error (10 %, variability of monitored start concentrations).*

l. 370: many teflon chambers are operated continuously under pressure to avoid contamination from outside. Did you determine the effectiveness of cleaning your chamber both with just flushing under pressure vs with evacuation? (same: L 400 → emptying it to 30% of its original volume). I wonder, what the impact of small leaks would be here. Did you perform leak tests, e.g. by spraying acetone outside the teflon chamber and monitoring it inside?

We would like to highlight the difference between the cleaning procedure and the experimental procedure. For cleaning (l. 370, original manuscript), we evacuated the chamber completely, thereby also actively sucking residual compounds from the walls. We then filled it with zero air and evacuated it again several times. As the blank values in the chamber were always very low (see Fig. 6), we are confident that our cleaning procedure was effective. The actual experiments were performed in batch mode. The statement about not emptying the chamber to less than 30% of its original volume (l. 400, original manuscript) refers to the experimental procedure. Not emptying the chamber completely had the purpose of (i) minimizing wall effects which become more complex when the area-volume-ratio increases, and (ii) avoiding contamination through leaks. We understand that the text was not perfectly clear about this. We therefore improved it by stressing which procedure was applied for cleaning vs for the experiments.

Old version (3.3.1, 1st paragraph):
Prior to each experiment, we cleaned the chamber and recorded blank values. To dilute and remove remaining impurities, the chamber was filled with zero air and then evacuated three times in total. Zero air was prepared by purifying pressurized synthetic air with a commercial zero air generator (Messer Griesheim GmbH, SL 50). In the first flushing cycle, the solar simulator was ignited to promote the oxidation of potential residuals and their release from the walls. After the cleaning, preliminary chamber blanks were measured by the SPME-GC-MS and the PTR-ToF-MS to confirm the cleanliness of the chamber and to test the instrumental performances.

New version (3.3.1, 1st paragraph):
Prior to each experiment, we cleaned the chamber and recorded blank values. To dilute and remove remaining impurities, the chamber was filled with zero air and then fully evacuated three times in total. Zero air was prepared by purifying pressurized synthetic air with a commercial zero air generator (Messer Griesheim GmbH, SL 50). In the first flushing cycle, the solar simulator was ignited to promote the oxidation of potential residuals and their release from the walls. After the cleaning, we filled the chamber with zero air again, and measured preliminary chamber blanks with the SPME-GC-MS and the PTR-ToF-MS to confirm the cleanliness of the chamber and to test the instrumental performances.

Old version (3.3.1, 4th paragraph):
The chamber was never depleted to less than 30 % of its initial volume to avoid too high surface-area-to-volume ratios and an increased importance of wall effects. After all the data were collected, the chamber was evacuated and then cleaned three times (zero air filling and subsequent evacuation).

New version (3.3.1, 4th paragraph):
During the experiments, the chamber was never emptied to less than 30 % of its initial volume to avoid (i) too high surface-area-to-volume ratios and an increased importance of wall effects as well as (ii) contamination through leaks. After all the data were collected, the chamber was completely evacuated and then cleaned three times (zero air filling and subsequent evacuation).

L 379 ff. – not completely clear to me. Is the toluene flushed with the N2 (15mins, 5slpm) into the larger teflon batch chamber or was the N2 flushing used to clean the toluene flask before toluene was injected? In the latter case, how is the toluene entering the batch chamber?

The first case, toluene was flushed with the N_2 into the chamber. We changed the text to make it clearer.

Old version (3.3.1, 2nd paragraph):
Upon completion of the preparatory work, we added the reagents sequentially. First, we introduced the VOC precursor into the chamber. The pure compound was injected through a septum into a 100 mL round-bottomed flask, which was flushed with N_2 as carrier gas (5 SLPM) for 15 minutes.

New version (3.3.1, 2nd paragraph):
Upon completion of the preparatory work, we added the reagents sequentially. First, we introduced the VOC precursor into the chamber. The pure compound was injected through a septum into a 100 mL round-bottomed flask, which was then flushed with N_2 as pick-up flow (5 SLPM) for 15 minutes.

fig. 3 – I suggest to add long names with the short forms in brackets in the figure description

We followed the suggestion, and added the short forms of the depicted toluene products to the figure caption.

Added text to the caption of Fig. 3:
[…] Toluene products in this figure include o-cresol (OCR), m-cresol (MCR), p-cresol (PCR), benzyl alcohol (BOH), and benzaldehyde (BAL).

For consistency, we added analogous statements to the captions of Fig. 10 and Fig. 11 as well.

Added text to the caption of Fig. 10:
[…] Toluene products in this figure include o-cresol (OCR), m-cresol (MCR), p-cresol (PCR), and benzyl alcohol (BOH).

Added text to the caption of Fig. 11:
[…] Toluene products in this figure include o-cresol (OCR), m-cresol (MCR), p-cresol (PCR), benzyl alcohol (BOH), and benzaldehyde (BAL).

We also added further explanation to the caption of Fig. S10.

Old version Fig. S10:
Time series of the two ISTDs and PFBHA as the carbonyl reagent are shown.

New version Fig. S10:
Time series of the two ISTDs (phenol-d_6 = PHE-d_6, and acetophenone-d_8 = APH-d_8) and PFBHA as the carbonyl reagent are shown.

l. 480 – You mention that the reactions with NO3 and O3 are negligible due to their small reaction rates. This is slightly confusing, as I thought the experiments were Nox-free and you did not mention that you actively added O3, so are the experiments not O3 and NO3-free?

We understand the confusion. For the O_3- and NO_3-related losses, we referred to the experiments in the presence of NO_x that we performed for method development but did not include in this manuscript to evaluate the toluene mechanism (see section 3.3: "In total, we performed 18 experiments at different temperatures and with different initial NO_x mixing ratios for method development purposes. To evaluate the product formation yields and gain mechanistic insight into the toluene chemistry, we focused here exclusively on six NO_x-free toluene-OH photooxidation experiments at T = 298±1 K and different degrees of photochemical aging."). We deleted this sentence to avoid confusion and since the experiments with added NO_x are not relevant to the mechanistic evaluation presented in this work.

Deleted text (end of Sect. 4):
Its rate constants for reactions with O_3 and NO_3 radicals are about $k < 10^{-21}$ molecules^{-1} cm^3 s^{-1} and $k = 7.8\times10^{-17}$ molecules^{-1} cm^3 s^{-1} (IUPAC, 2024), so that these loss processes were irrelevant under all experimental conditions.

Fig. 4: the legend could be improved for more clarity. There are e.g. multiple grey lines in different shades of gray, but only 1 line in the legend. Alternatively, decreasing the visibility of the error-minima and maxima traces and just using shading might help

As suggested, we decreased the visibility of the error minima and maxima and just used shading to improve clarity. We also included the mercury emission lines with a distinct line pattern in the legend.

To improve comprehensibility, we furthermore labelled the panels, and referred to them in the figure caption and the main text accordingly.

[see Figure 4]

Old version caption Fig. 4:
[...] The left figure shows the emission spectrum of the solar simulator (7 HMI lamps with bandpass filter) available from [...]. In the right figure, the solar simulator spectrum is compared against the natural actinic flux as calculated [...]

New version caption Fig. 4:
[...] (a) The emission spectrum of the solar simulator (7 HMI lamps with bandpass filter) available from [...]. (b) The solar simulator spectrum compared against the natural actinic flux as calculated [...].

Old version main text (sect. 4.1.1):
The measured emission spectrum of the solar simulator is shown in Fig. 4. [...] For comparison, Fig. 4 shows the resemblance between the solar simulator emission spectrum and the natural solar spectrum, [...]

New version main text (sect. 4.1.1):
The measured emission spectrum of the solar simulator is shown in Fig. 4a. [...] For comparison, Fig. 4b shows the resemblance between the solar simulator emission spectrum and the natural solar spectrum, [...]

l. 523: how did you find the OH concentration? Was it measured via the toluene decay?

Yes, we calculated the OH radical concentration via the decay of toluene. To make this clearer, we added more explanation to the calculation of the OH exposure in the section about the loss corrections. We also specifically added the calculation of the OH radical concentrations.

Old version (end of sect. 3.3.3):
To treat the reaction with OH radicals using the same correction formula, we converted the second order reaction of any of the analytes with OH radicals to a pseudo first order approach, as shown in Eq. (3):

$$\Delta t \times k' = \Delta t \times k \times [OH] = \Delta OHexp \times k \quad (3)$$

where k' is the pseudo first order rate constant in s^{-1}, k is the second order rate constant in molecules^{-1} cm^3 s^{-1}, $[OH]$ is the OH radical concentration in molecules cm^{-3}, and $\Delta OHexp$ is the difference of the OH exposure between the two data points which is given in the units of molecules cm^{-3} s. The second order rate constants for the reactions of the photooxidation products with OH radicals were obtained from the IUPAC recommendations and the scientific literature (Table S7). The calculation of the OH exposure at a given point in time was based on the observed decay of toluene to circumvent the need for direct OH radical measurements, as shown in Eq. (4):

$$\Delta OHexp = OHexp_t - OHexp_{t-1} \quad (4a)$$

$$OHexp_t = ln\left(\frac{[toluene]_t}{[toluene]_0}\right) / -k_{toluene+OH} \quad (4b)$$

where $OHexp_t$ and $OHexp_{t-1}$ are the OH exposures determined for the time steps associated with the target data point and the previous data point in molecules cm^{-3} s, $[toluene]_t$ and $[toluene]_0$ are the concentrations of toluene at the given point in time and at the beginning of the experiment in molecules cm^{-3}, and $k_{toluene+OH}$ = 5.6×10^{-12} molecules^{-1} cm^3 s^{-1} is the well-characterized second order rate constant of toluene with OH radicals (IUPAC, 2024).

New version (end of sect. 3.3.3):
To treat the reaction with OH radicals using the same correction formula, we converted the second order reaction of any of the analytes with OH radicals to a pseudo first order approach. For this purpose, we referred to the OH exposure as the time-integrated OH radical concentration, as shown in Eq. (4):

$$OHexp_t = \int_{t=0}^{t=t} [OH] \Delta t \quad (4a)$$

$$\Delta OHexp = \int_{t=t-1}^{t=t} [OH] \Delta t \quad (4b)$$

$$\Delta t \times k' = \Delta t \times k \times [OH] = \Delta OHexp \times k \quad (4c)$$

where $OHexp_t$ is the OH exposure between time t and the beginning of the experiment in units of molecules cm^{-3} s, $[OH]$ is the OH radical concentration in molecules cm^{-3}, $\Delta OHexp$ is the difference of the OH exposure between the two data points in molecules cm^{-3} s, k' is the pseudo first order rate constant in s^{-1}, and k is the second order rate constant in molecules^{-1} cm^3 s^{-1}. The second order rate constants for the reactions of the photooxidation products with OH radicals were obtained from the IUPAC recommendations and the scientific literature (Table S7). To

circumvent the need for direct OH radical measurements, the OH exposure at a given point in time was calculated based on the observed decay of toluene, as shown in Eq. (5):

$$\Delta OHexp = OHexp_t - OHexp_{t-1} \quad (5a)$$

$$OHexp_t = ln\left(\frac{[toluene]_t}{[toluene]_0}\right) / -k_{toluene+OH} \quad (5b)$$

where $OHexp_t$ and $OHexp_{t-1}$ are the OH exposures determined for the time steps associated with the target data point and the previous data point in molecules cm^{-3} s, $[toluene]_t$ and $[toluene]_0$ are the concentrations of toluene at the given point in time and at the beginning of the experiment in molecules cm^{-3}, and $k_{toluene+OH}$ = 5.6×10^{-12} molecules^{-1} cm^3 s^{-1} is the well-characterized second order rate constant of toluene with OH radicals (IUPAC, 2024). Based on the temporally resolved OH exposure (available in minute resolution according to the toluene data set), we used the relationship in Eq. (4b) to calculate OH radical concentrations averaged over the time intervals between t and t-1 according to Eq. (6):

$$[OH] = \Delta OHexp / \Delta t \quad (6)$$

Since we added equation 6, we needed to change the numbering of the following equations accordingly.

[see following equations]

L. 525: that's good! But I believe, this needs to be checked before future experiments after NOx has been added, again.

We agree. As reviewer 1 also found the statement concerning HONO too speculative, we deleted this part entirely, so as to avoid false implications.

Old version (Sect. 4.1.1, OH radical paragraph):
We did not observe any product formation in a test run without H_2O_2 addition, confirming H_2O_2 photolysis as the main source of OH radicals, rather than for example the release of HONO from chamber walls which would produce both OH radicals and NO.

New version (Sect. 4.1.1, OH radical paragraph):
We did not observe any product formation in a test run without H_2O_2 addition, confirming H_2O_2 photolysis as the main source of OH radicals.

Fig. 7: “relative response” is not clear from the figure or figure description (only after reading the text). It is not described, in relation to what the signals are shown. Please add units for more clarity.

We agree that the notion “relative response” is not clear just from the figure. We added the information “area analyte / area internal standard” to the y axis label. We also added the unit of the slope m (“molecules^{-1} cm^{3}”).

[see Figure 7]

Fig. 8: please add the shortforms (e.g. PHE-d6) in the figure caption

We added the shortforms in the figure caption as suggested.

Original version caption Fig. 8:
Each individual marker represents one measurement of phenol-d_6 or acetophenone-d_8, respectively.

New version caption Fig. 8:
Each individual marker represents one measurement of phenol-d_6 (PHE-d_6) or acetophenone-d_8 (APH-d_8), respectively.

Fig. 11: instead of "final concentration", maybe call it total produced concentration or similar? And mark it in the plot as a horizontal bar with uncertainties (cause the loss rates also have uncertainties)

We agree that the term was imprecise. We replaced "final concentration" with "loss-corrected concentration", as we used this term throughout the paper.

Change in the main text:

Old version (beginning Sect. 4.3.1):
The contribution of the three characterized loss processes (wall losses, photolysis, secondary OH radical reactions) to the final concentration differed [...]

New version (beginning Sect. 4.3.1):
The contribution of the three characterized loss processes (wall losses, photolysis, secondary OH radical reactions) to the loss-corrected concentration differed [...]

Change in the caption of Fig. 11:

Old version:
Contributions of the non-corrected data and each of the three included loss corrections (wall losses, photolysis, secondary OH radical reactions) to the final concentration. [...]

New version:
Contributions of the non-corrected data and each of the three included loss corrections (wall losses, photolysis, secondary OH radical reactions) to the loss-corrected concentration. [...]

As suggested, we also specifically plotted the loss-corrected mixing ratios and the associated uncertainties in Figure 11. We furthermore labelled panel a and b to increase comprehensibility and adjusted the caption accordingly.

[see Figure 11]

Old version of caption:
[...] The shares of the raw data and the losses are shown on an absolute scale (mixing ratios, upper graph) and on a relative scale (normalized to the total corrected mixing ratio, lower graph). [...].

New version of caption:
[...] The shares of the raw data and the losses are shown on (a) an absolute scale (mixing ratios) and (b) on a relative scale (normalized to the loss-corrected mixing ratio). [...]

Fig. 12: please just write mixing ratio instead of the new shortform “MR”

We followed the suggestion and used the full notation “mixing ratio”. For consistency, we made the same adjustment in Fig. 13.

Please note that we also fixed a mistake in Fig. 12. Previously, two lines were shifted to a wrong position during the rendering of the plot.

[see Figures 12 and 13]

fig. S6.4: is m your calibration factor? Y-axis unit missing...

We added the unit “area” to the y axis label. We also added the unit of the slope m (“area molecules^{-1} cm^{3}”).

[see Figure S11 in Section S6.4]

Further note

As part of this manuscript, we present a weighted calibration for the mass *m/z* 109.0626 at the PTR-ToF-MS, assigned to the chemical structure C_7H_8O. In the original preprint, we considered the three cresol isomers as well as benzyl alcohol. During the review of the preprint, we improved the assignment. As the PTR-ToF-MS does not show any signal for benzyl alcohol at this mass due to fragmentation, we decided that the inclusion of benzyl alcohol within the weighted calibration can be misleading. In particular, the difference between the averaged and weighted calibration factor is disproportionately large when including zero-sensitivity compounds. Therefore, we suggest the following modifications:

- to change the calculation and interpretation of the weighted PTR-ToF-MS calibration to include only the three cresol isomers, using their individual sensitivities and relative abundances (0.74:0.04:0.22 for the *o/m/p* isomer, respectively).
- to compare the sum signal at *m/z* 109.0626 against the sum of the three cresol isomers (no longer including benzyl alcohol) at the SPME-GC-MS accordingly (Fig. 9).
- to not evaluate and show any PTR-ToF-MS time series of benzyl alcohol (Fig. 10), as (i) it is not measurable at *m/z* 109.0626 and would no longer be represented in the weighted calibration, and (ii) we prefer not to analyse the unspecific *m/z 91.0522* fragment.

While the PTR-ToF-MS cresol data is not a major part of the manuscript, it is nevertheless addressed in a few different sections. We suggest to make all necessary changes for consistency, as outlined in the following:

Abstract:

Old version:
For the cresols and benzyl alcohol, we compiled a weighted calibration factor for the PTR-ToF-MS, taking into account isomer-specific sensitivities as well as the relative distribution as determined by the SPME-GC-MS. The weighted calibration improved the instrumental agreement to 15 %, whereas the PTR-ToF-MS overestimated the sum of the isomers by 25 % compared to the SPME-GC-MS concentrations when using the averaged calibration factor.

New version:
For the cresols, we compiled a weighted calibration factor for the PTR-ToF-MS, taking into account isomer-specific sensitivities as well as the relative distribution as determined by the SPME-GC-MS. The weighted calibration improved the instrumental agreement to 14 %, whereas the PTR-ToF-MS overestimated the sum of the isomers by 31 % compared to the SPME-GC-MS concentrations when using the averaged calibration factor.

Methods (last paragraph 3.2.2):

Old version:
[…] All compounds were analysed at the protonated mass of m/z 109.0626, yet the instrumental response of each of the isomers was derived individually. For analyzing the sum signal during the photooxidation experiments, we calculated the weighted sensitivity for each of the isomers as the product of its instrumental sensitivity and its relative abundance. The relative abundances

were derived as fixed values from the formation yields as determined with the SPME-GC-MS data (Sect. 4.3.2). [...] From the correctly quantified sum signal, we later extracted the experimental concentrations of the individual C_7H_8O isomers using the same fixed relative abundances.

New version:
[...] All compounds were analysed at the protonated mass of m/z 109.0626, yet the instrumental response of each of the isomers was derived individually. For analyzing the sum signal during the photooxidation experiments, we calculated the weighted sensitivity for each of the isomers as the product of its instrumental sensitivity and its relative abundance. Since benzyl alcohol did not have any signal at this mass (Table S6), only the cresols isomers were included in the weighted calibration. Their relative abundances were derived as fixed values from the formation yields as determined with the SPME-GC-MS data (Sect. 4.3.2). [...] From the correctly quantified sum signal, we later extracted the experimental concentrations of the individual cresol isomers using the same fixed relative abundances.

<u>Results (section 4.2.3):</u>

- 2nd paragraph:

Old version 2nd paragraph:
[...] Benzyl alcohol did not show any signal on m/z 109.0626 and thus did not contribute to the sum signal. Instead, it fragmented to m/z 91.0522, which is in accordance with the literature (Yeoman et al., 2021) and probably indicates the formation of the benzyl radical after abstraction of water from the protonated molecule. The isomer-specific calibration was performed with relative abundances of 0.71 for o-cresol, 0.04 for m-cresol, 0.21 for p-cresol, and 0.04 for benzyl alcohol (Sect. 4.3.2). As o-cresol is by far the most abundant isomer, its contribution to the weighted calibration factor on m/z 109.0626 was as high as 87 %. Overall, considering the relative abundances and the specific sensitivities of the individual isomers led to a 1.47-fold increase of the calibration factor compared to the non-weighted calibration (Table S6).

New version 2nd paragraph:
[...] Benzyl alcohol did not show any signal on m/z 109.0626 and thus did not contribute to the sum signal and was excluded from the weighted calibration. Instead, it fragmented to m/z 91.0522, which is in accordance with the literature (Yeoman et al., 2021) and probably indicates the formation of the benzyl radical after abstraction of water from the protonated molecule. As this fragment is unspecific, we did not evaluate it further. The isomer-specific calibration was performed with relative abundances of 0.74 for o-cresol, 0.04 for m-cresol, and 0.22 for p-cresol (Sect. 4.3.2). As o-cresol is by far the most abundant isomer, its contribution to the weighted calibration factor on m/z 109.0626 was as high as 87 %. Overall, considering the relative abundances and the specific sensitivities of the individual isomers led to a 1.15-fold increase of the calibration factor compared to the non-weighted calibration (Table S6).

- Figure 9:

In Figure 9, the first panel now shows a comparison of the sum of the three cresol isomers, not of all C_7H_8O isomers, changing both the x and y data. The caption was adapted accordingly.

[See Fig 9]

Old caption:
[…] In case of the cresol isomers and benzyl alcohol, the non-weighted and weighted calibrations for the sum of the isomers are distinguished. Error bars represent the quantification error of the respective instruments. Error bars of the sum of the C_7H_8O isomers (cresols and benzyl alcohol) at the SPME-GC-MS are the weighted sum of the individual errors. […]

New caption:
[…] In case of the cresol isomers, the non-weighted and weighted calibrations for the sum of the isomers are distinguished. Error bars represent the quantification error of the respective instruments. Error bars of the sum of the cresol isomers at the SPME-GC-MS are the weighted sum of the individual errors. […]

- 3rd paragraph:

Old version 3rd paragraph:
[…] For the cresol isomers and benzyl alcohol, the weighted calibration improved the coherence between the two instruments from 25 % to 15 % deviation. […]

New version 3rd paragraph:
[…] For the cresol isomers, the weighted calibration improved the coherence between the two instruments from 31 % to 14 % deviation. […]

- 4th paragraph:

Old version 4th paragraph:
[…] Rather than performing a separate weighted calibration for each time segment assigned to a specific SPME-GC-MS sample, we used a fixed relative abundance of the C_7H_8O isomers for both the weighted calibration and the experimental data. This is a reasonable assumption, as their sinks in the BATCH Teflon chamber (wall losses, photolysis, secondary OH radical reactions) agree within 26 % (Tables S1, S7, S8). The experimental SPME-GC-MS data confirm that the variability of the isomeric distribution over the course of the conducted experiments is only 17 % and hence within the quantification error. In addition to the simplicity of implementation, the adoption of a fixed relative abundance also makes the PTR-ToF-MS results less susceptible to the random error of the SPME-GC-MS data.

New version 4th paragraph:
[…] Rather than performing a separate weighted calibration for each time segment assigned to a specific SPME-GC-MS sample, we used a fixed relative abundance of the cresol isomers for both the weighted calibration and the experimental data. This is a reasonable assumption, as their sinks in the BATCH Teflon chamber (wall losses, photolysis, secondary OH radical reactions) agree within 16 % (Tables S1, S7, S8). Accordingly, the isomeric distribution derived from the formation yields (including loss corrections) is equal to the average isomeric distribution calculated from the measured concentrations (excluding loss corrections). The variability of the isomeric distribution in the experimental SPME-GC-MS data is only 13 % and hence within the quantification error. In addition to the simplicity of implementation, the adoption of a fixed relative abundance also makes the PTR-ToF-MS results less susceptible to outliers and the random error of the SPME-GC-MS data.

- Figure 10:

In Figure 10, benzyl alcohol was excluded. The caption was adapted accordingly.

[See Fig 10]

Old version caption:
Figure 10: Time series of the cresol isomers and benzyl alcohol by SPME-GC-MS and PTR-ToF-MS. [...] In the depicted experiment, the sum of the isomers agrees exceptionally well between the two instruments, while the measured isomeric distribution deviates slightly from the fixed relative abundance. Toluene products in this figure include o-cresol (OCR), m-cresol (MCR), p-cresol (PCR), and benzyl alcohol (BOH).

New version caption:
Figure 10: Time series of the cresol isomers by SPME-GC-MS and PTR-ToF-MS. [...] Toluene products in this figure include o-cresol (OCR), m-cresol (MCR), and p-cresol (PCR).

<u>Conclusion:</u>

Old version:
We calibrated the main first generation products of toluene for both instruments, and found good agreement for the ring-retaining products (the cresol isomers, benzyl alcohol, benzaldehyde). We further demonstrated that the sensitivity of the PTR-ToF-MS for different structural isomers can vary substantially. For the cresols and benzyl alcohol, we therefore applied a weighted calibration factor based on the determined isomer-specific sensitivities and the distribution of the isomers as obtained by the SPME-GC-MS data. This led to an improved instrumental agreement, and [...]

New version:
We calibrated the main first generation products of toluene for both instruments, and found good agreement for the cresol isomers and benzaldehyde. We further demonstrated that the sensitivity of the PTR-ToF-MS for different structural isomers can vary substantially. For the cresols, we therefore applied a weighted calibration factor based on the determined isomer-specific sensitivities and the distribution of the isomers as obtained by the SPME-GC-MS data. This led to an improved instrumental agreement, and [...]

<u>Supplement:</u>

- Table S6

In the table, we differentiated the cresol isomers and benzyl alcohol more explicitly, showing plainly that benzyl alcohol is not incorporated into the weighted calibration. We still list its zero sensitivity at *m/z* 109.0626, as we believe this to be of interest to the community. For the weighted calibration, we updated all the values. Note that also the non-weighted values now appear different, as we performed the isomer-specific calibrations at a slightly later point and scaled them to the previously recorded sum signal, which is now composed of three instead of

four compounds. We also changed the caption of the table, specifying the relative abundances as well as changing the mentions of the C_7H_8O isomers to only the cresol isomers.

[see Table S6]

Old version caption:
[…] For the C_7H_8O isomers, we derived a weighted sensitivity by multiplying the recorded sensitivity with the relative abundance as determined by SPME-GC-MS. The weighted calibration factor for the sum signal was obtained from the sum of these weighted isomer-specific sensitivities. When evaluating individual C_7H_8O isomers, we adopted the error and LOD of the weighted sum of the isomers.

New version caption:
[…] For the cresol isomers, we derived a weighted sensitivity by multiplying the recorded sensitivity with the relative abundance as determined by SPME-GC-MS (0.74 for o-cresol, 0.04 for m-cresol, and 0.22 for p-cresol). The weighted calibration factor for the sum signal was obtained from the sum of these weighted isomer-specific sensitivities. When evaluating individual cresol isomers, we adopted the error and LOD of the weighted sum of the isomers.

- Section S7

In Section S7, we changed the wording in the introductory text and the caption of Figure S14, and updated the displayed PTR-MS data in Figure S14.

[see Fig. S14]

Old version text:
[…] for o-cresol as one of the C_7H_8O isomers […]

New version text:
[…] for o-cresol as one of the cresol isomers […]

Old version caption:
[…] The mixing ratio of o-cresol at the PTR-ToF-MS was calculated from the weighted calibration and the fixed relative abundance of the C_7H_8O isomers (0.71 for o-cresol).

New version caption:
[…] The mixing ratio of o-cresol at the PTR-ToF-MS was calculated from the weighted calibration and the fixed relative abundance of 0.74 for o-cresol.