We thank the referees for their insightful questions and comments, which helped improving the quality of the paper. Our answers to referee #2 are listed in the following in red, after the reviewer's comments, which are in black. The modifications in the text are marked in yellow.

Anonymous Referee #2

This manuscript described the design and characterization of an online ROS analysis systems very similar to instruments built by other groups before. A number of aspects are not described in sufficient detail and the following points need to be considered before publication.

p.1, line 14/15 (Abstract): it is unclear what the detection limit for offline is. 1.3 or 9-13nmol L-1. Also, indicate with what compound this detection limit was determined. nmol H2O2 L-1?

We report two types of detection limits, the instrument detection limit and the method detection limit. The details of the definition and their determination are given in Sect. 3.1.1. The instrument LOD was 1.3 nmol L^{-1}, determined as three times the standard deviation of the background when OF-UPW was injected into the sampling line. The method LOD was determined based on the reproducibility of the instrument background and the filter blanks. Due to the varying background of the offline instrument between different samples and the variation in filter blanks and filter extraction a higher method detection limit was found. Due to another question raised by this referee we re-evaluated the method detection limit (see below). Instead of 9-13 nmol L^{-1} we give now 18 nmol L^{-1} as the highest determined method detection limit. Our instrument is calibrated with H_2O_2.
We added in the abstract that the instrument is calibrated with H_2O_2.

p.2, line 15: Please also reference Wang et al., Journal of Toxicology, 2011, who first developed an online DCFH system.

We added this reference.

p.3, line 12: What is the difference between reaction and incubation?

We used it in the way that incubation relates to the experimental conditions and reaction to the chemical process under those conditions. As incubation is a well-defined process in biology and medicine and does not conform to our use, we changed the terminology. We replace incubation by reaction throughout the manuscript.

p.3, line 15/16: For how long was the denuder efficient and how was this assessed? The HRP assay is sensitive to H2O2. Was the denuder efficient in removing gaseous H2O2? How was that determined?
We tested the denuder with two gaseous oxidants, ozone and NO_2. We did not observe an increase in signal even at a dose of about 500 ppb over 5 hours. H_2O_2 mixing ratios in ambient air are at least two orders of magnitude lower. Therefore, we expect no breakthrough of H_2O_2. To be on the safe side we exchanged the denuder for each laboratory experiment (~ 5 h). Further, we checked the ROS blank by measuring the particle-free air (through the particle filter and the denuder) to see if the ROS blank signal increases due to gas phase break through. This was only observed in some wood burning experiments with extremely high pollutant concentrations.
We added in Sect. 3.2 on gas phase interference tests: Based on these results we assume that gaseous H_2O_2 is also completely removed.

We also mention in the text how long the denuder was used (page 7): 2) replacing the hydrophilic and hydrophobic filters in the aerosol collector and the denuder every 2-3 days during ambient measurements; while in laboratory experiments, we exchanged the denuder for each laboratory experiment (~ 5 h) to be on the safe side;

p.4, line 12/13. Please support the statement in this sentence with evidence. By how much was the lifespan of the solution shortened and how did the additional contamination affect the measurement?

We tested the auto-oxidation of the working solution containing both HRP and DCFH in the same way as in King and Weber (2013). By mixing only OF-UPW with the HRP-DCFH working solution the signal, which is actually the background, increased with a rate of 0.9%/h. This means that there is a slow reaction with the dissolved oxygen consuming the DCFH and shortening the lifespan of the HRP-DCFH solution.

When the sample is extracted online with the HRP solution as in Fuller et al. (2014), the HRP needs to go through the aerosol collector, where contaminants adsorbed on the hydrophilic/hydrophobic filters or the oxygen in the mist chamber might react with HRP and then oxidize DCFH as described by Berglund et al. (2002) and modified by Fuller et al. (2014). We added this information in the modified manuscript (Page 4): We tested the auto-oxidation of the working solution containing both HRP and DCFH. By mixing only OF-UPW with the HRP-DCFH working solution the signal, which is actually the background, increased with a rate of 0.9%/h. This means that there is a slow reaction with the dissolved oxygen consuming the DCFH consequently shortening the lifespan of the HRP-DCFH solution. When the sample is extracted online with the HRP solution as in Fuller et al. (2014), the HRP needs to go through the aerosol collector, where contaminants adsorbed on the hydrophilic/hydrophobic filters or the oxygen in the mist chamber might react with HRP and then oxidize DCFH as described by Berglund et al. (2002) and modified by Fuller et al. (2014). Therefore, we used only OF-UPW to extract the aerosol samples. The DCFH and HRP reagents were kept separate and were only mixed together right before the aerosol aqueous extract was added.

p. 7, section 2.2 I am not sure this section is necessary as it does not add any information.

We decided to add this section because we have tested many aspects and it might be hard for a reader to quickly grasp all the purpose and the extent of those tests. We think this Section is useful to guide the reader and we would like to keep it.

p.8, line 1: With what experiment was the residence and response time determined.

The residence time was determined from the time of injection of an H_2O_2 solution to the time the fluorescence signal started to increase. The response time was determined as the time required for the signal to increase from 10% to 90% of its final value. We added this information in the manuscript on page 8: The residence and response time of the sample in the instrument were measured to be approximately 19 min and 8 min, respectively. The former was determined as the time from injection of an H_2O_2 solution to the time the fluorescence signal started to increase while the response time corresponds to the rise time of the fluorescence signal from 10 % to 90 % of the full signal.

Fig. 2: How is the difference in detector response reconciled between the two compounds?

In Fig. 2 we discuss the influence of the reaction time on the detector response of these two compounds H_2O_2 and 2-butanone peroxides. Doubling of the reaction time did not increase the signal strength for both compounds. The difference in detector response of various ROS is shown in Sect. 3.1.2 & Fig. 3. Many of the tested ROS have a lower response compared to H_2O_2. This has probably to do with the peroxidase, which mediates the reaction of the peroxide with DCFH. However, the mechanism of this is not known and would need detailed investigations, which are outside of the scope of this paper.

At the end of the discussion (page 11) we added the following sentence: Thus, we regard the ROS signal measured by the DCFH assay as a lower limit for the effective ROS content.

p.8, line 8: Was the LOD online determined by using ambient air?
Yes, the online LOD was determined by measuring ambient air passing through the denuder and the particle filter. We clarified this in the modified manuscript (page 9): Under normal instrument operation condition, an instrument limit of detection (LOD) of 2 nmol m^{-3} of sampled ambient air was determined for the online methodology.

p.9 & Fig.S3: Looking at the data in Fig.S3 and confidence intervals shown it look to me the detection limit is more in the range of 15 nM. How does that compare with numbers discussed on p.8.

We assume that you calculated the 15 nM from the confidence interval at very low H_2O_2 concentrations. We determined the method detection limit by two approaches: 1) injecting several times different batches of OF-UPW (LOD = 9 nM); 2) extraction of different blank filters (LOD = 13 nM). If we calculate an LOD from the calibration curve of Fig. S3 we obtain 18 nM. This LOD would include both the variability of the background (OF_UPW and blank filter) and the conditions of the extraction. We decided to take the largest of these three numbers (18 nM) as the method detection limit. This discussion is added in the manuscript (page 10): Based on this, the uncertainty of H_2O_2 at extremely low concentrations would be 18 nM. This is larger than the method LOD determined above from the OF-UPW and blank filters. We consider the largest of these uncertainties (i.e.18 nM) as our final method LOD.

p.9, line 20: Did the use of ethyl acetate affect the reactivity of HRP and/or DCFH? How was this verified? The enzyme HRP could be strongly impaired in its reactivity in an organic solvent.

Studies have found that only about 10 % of the initial activity of HRP was lost after incubation in ethyl acetate (EA) for 12 hours. The influence of EA on the HRP activity is much lower than by other organic solvents like acetone, tetrahydrofuran (Jang et al., 2014). To prepare the calibration solutions we first dissolved the pure chemicals in ethyl acetate (EA) and then diluted this solution ~10000 times in water to reach the desired concentration. This means that the EA concentration is roughly 1 mmol/L when mixed with HRP/DCFH. We believe that its influence on HRP/DCFH is negligible at such a low concentration.

p. 10, line 5: Data are only shown from 30-150nM, not 0 – 150nM.
We have added the 0 nM measurements in Fig. 3.

p. 11, Fig. 4: No detail is given for the data shown in Fig. 4. What are the dates, collected? Do they correspond to data shown in the references cited? Why is there a difference in the two data sets, what are the errors on the data shown etc.
Indeed, the description of the sites and filter sampling methods in Milan and San Vittore is given in the references cited. We mention this now (page 12): More details on the analysis of the samples can be found in the cited references. The different slopes between these two data sets might be due to the different emission sources (traffic in Milan and wood combustion in San Vittore) at these two locations as investigated in the cited papers (Perrone et al., 2016; Zotter et al., 2014). We add this information in the modified manuscript (page 12): The different slopes between these two data sets might be due to the different emission sources (traffic in Milan and wood combustion in San Vittore) at these two locations (see Perrone et al., 2016; Zotter et al., 2014).

We also added error bars on the data representing our measurement precision (Fig. 4 & legend).

p.13, line 14: SO4 and NO3 (given in units of ug m-3) were mixed with H2O2 (given in units of nM) for cross sensitivity test. How was that done. Was the sulfate and nitrate nebulised as aerosol? A lot more detail needs to added here.

We prepared solutions of ~ 1.38 µM SO_4^{2-} and ~20 µM NO_3^- from $(NH_4)_2SO_4$ and NH_4NO_3, respectively. Such concentrations would typically be received after collection of 23.5 µg m^{-3} SO_4^{2-} and of 228 µg m^{-3} NO_3^- with the online instrument. This is equivalent to 5 and 30 times higher concentrations than the ambient concentration as observed in Milan (Perrone et al., 2016). The measurement of these concentrations is then compared to the mixture of similar concentrations with 115 nM H_2O_2 and 272.5 nM 2-BP, as listed in Table 2. To make it more clear, we modified the text and added the prepared concentration of the SO_4^{2-} and NO_3^- solutions (page 14-15): Therefore, we tested the fluorescence response to ~ 1.38 µM SO_4^{2-} and ~20 µM NO_3^- solutions prepared from $(NH_4)_2SO_4$ and NH_4NO_3, respectively. Such concentrations would typically be observed after collection of 23.5 µg m^{-3} of SO_4^{2-} and 228 µg m^{-3} of NO_3^- with the online instrument. This is equivalent to ~ 5 and ~ 30 times higher concentrations than observed in Milan (Perrone et al., 2016). These measurements are then compared to cross sensitivity tests of ~1.38 µM SO_4^{2-} and ~20 µM NO_3^- with 115 nM H_2O_2 and 272.5 nM 2-BP (Table 2).

p. 14/Fig. 6: The experiments with Fe2+/Fe3+ are discussed in a purely descriptive way. A more detail discussion rationalising the results and referencing Fenton reactions is needed.

We added the reaction equations to the description and give some more explanations. It reads now (page 15):

This might be due to the consumption of a substantial amount of H_2O_2 by Fe^{2+}, for the production of $HO\cdot$ ($Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + HO\cdot$), which will further react with H_2O_2 and resulting in further reduction of the H_2O_2 concentration ($HO\cdot + H_2O_2 \rightarrow H_2O + HO_2\cdot$; $HO_2\cdot + H_2O_2 \rightarrow O_2 + H_2O + HO\cdot$)(Kolthoff and Medalia, 1949). This indicates that concentrations of soluble $Fe^{2+} \leq 1$ nM, which were obtained at ambient concentrations of ≤ 10 ng m^{-3} soluble Fe^{2+} in the online instrument, will not influence the ROS measurement. However, in cases of high ambient soluble Fe^{2+} concentrations the ROS signal might be reduced, whereby this also depends on the H_2O_2 equivalent concentration. Measured ambient iron concentrations were found to be in the range of tens to several thousands of ng m^{-3} (Perrone et al., 2016; Oakes et al., 2012; Visser et al., 2015). Oakes et al. (2012) reported that water soluble Fe (II) constitutes between 2.5 and 32 % of total iron, resulting in a water soluble Fe (II) concentration up to 30 ng m^{-3}, which would be equivalent to ~ 2 nM in our online instrument. According to our 1st pair of experiments in Fig. 6 (1 nM Fe^{2+} mixed with 113 nM H_2O_2 solution) this would not suppress the ROS signal.

p.14, line 11: It is mentioned that water-soluble Fe2+ is measured up to 100s ng m-3 in ambient samples. To what concentration does that correspond in the working solution. Could the ROS signal potentially be suppressed under these conditions? This should be discussed more clearly.
Oakes et al. (2012) report water-soluble Fe^{2+} up to 30 ng/m^3, which is equivalent to ~ 2 nM in the online instrument. This concentration is too small to suppress the ROS signal according to our 1st pair of experiments (where we prepared 1 nM Fe^{2+} mixed with 113 nM H_2O_2 solution). To further confirm this, we measured the influence of genuine atmospheric particulate metals on the ROS signal in Fig. 7 and the results showed no influence either. We explain this now in the modified version (see answer above).

p. 17/Fig. 8: Are the units for the x and y axis "nmol H2O2 m-3"? If yes, this should be indicated explicitly.
Yes, they are. We modified the units in Figure 8 accordingly.

p.16, line 15/Fig.8: It is mentioned that filters were collected before and after the VACES. Was ROS different in the filters collected before and after the VACES? Did the use of a VACES affect online ROS concentrations?

We did not collect filters before and after VACES at the same time. Thus, a direct comparison cannot be done. We compared the filters collected before and after the VACES with the on-line measurements, which were always done after the VACES. No systematic trend in concentrations was observed. We added this information in the new version of the manuscript (page 18): We did not observe a systematic difference between ROS concentrations on filters taken before and after VACES compared with the online measurements.
During reanalysis of this data, we have noticed that the ROS didn't response to the short changes enrichment factor during the adjustment of the setup, thus we readjusted the calculation of enrichment factors for these periods. This improvement led to minor changes in the concentrations of the online measurement (Fig. 8), but did not affect any of the conclusions.

Fig 9 and related discussion: It is not acceptable to derive a half-life of ROS from the data shown in Fig. 9. The time resolution of the data presented is far too sparse to constrain the half-life of ROS to any reasonable accuracy. This Figure has to be deleted or much more data has to be provided to make a meaningful statement about ROS half-life.

We think the reviewer has raised a valid point. Therefore, in the corrected version of the manuscript we have reconsidered the fitting procedure to carefully account for the uncertainties. The obtained parameters are statistically not different from those obtained before (differences within 10 %) and the related uncertainties and additional limitations of the approach used are now discussed in the corrected version of the manuscript in Sect. 3.4.2 (page 20):

The measured ROS concentrations in SOA from the different wood combustion experiments exhibit a clear decrease with increasing filter storage duration (Fig. 9). In addition, this decay seems to follow a double exponential function. This indicates the presence of a short-lived fraction A_1 with a decay constant $\pi_1 = \ln(2)/T_1$ and a long-lived fraction A_2 with a decay constant $\pi_2 = \ln(2)/T_2$, where T_i represents the half-life. A biexponential decay function was applied to fit the experimental values, whereby the two decay constants are considered to be the same for all experiments:

$$ROS_{norm}(t) = A_1 * EXP(-\pi_1 * (t - t_1)) + A_2 * EXP(-\pi_2 * (t - t_1))$$ Eq. (3)

Here $A_{2,i} = 1 - A_{1,i}$, $0 \leq A_{1,i}$, $A_{2,i} \leq 1$, i refers to an experiment number, t is the time after sampling and t_1 is the time when the first offline measurement was performed. $ROS_{norm}(t)$ is the ROS measured at time t normalized to the ROS measured at time t_1. The model parameters and their respective uncertainties are shown in Table 3. Measured and modeled values are compared in Fig. 9.

The results show that the two ROS fractions have highly different reactivity. The final modelling yields $\pi_1 = 9.68 \pm 2.56$ and $\pi_2 = 0.0016 \pm 0.0019$. The second fraction (long-lived) appears to be not reactive within our uncertainties and experimental time scales, as the associated reaction rate, π_2, is not statistically different from 0. The first fraction (short-lived) is highly reactive, with a half-life time $T_1 \approx 1.7 \pm 0.4$ h, similar reaction time-scales and extents were observed for SOA from α-pinene ozonolysis (Krapf et al., 2016). The uncertainty analysis suggests that we are capable of determining the reaction rate of reactive ROS, but not that of the long-lived ROS. The fraction of the long-lived ROS ($A_{2,i}$) could be determined with acceptable errors of 20 %. The main aim of the model is to show that the fraction of unstable ROS may vary significantly between experiments, but could be as high as 75 %, which highlights the need for an online ROS measurement technique. This variability in the

contribution of the unstable ROS fraction could be related to the burning conditions, in this study (shown in Figure S5).

We also discussed the main limitations of the model and this part reads as follows (page 20-21):

The model considers ROS to be composed of two components with different decay rates. However, we do expect that the OA contains spectrum of ROS with a wide range of reactivities. The model is thus a simplification of the ROS in the aerosol. Another simplification is that the decay rates of these two ROS components are considered to be the same between experiments. This may explain the reasons behind the high uncertainties in determining the rates, but does not have a significant effect on the determination of the contributions of the two fractions, $A_{1,i}$ and $A_{2,i}$. We also note that the decay rates and the ROS fractions determined from our results are specific for biomass burning SOA and cannot be extrapolated to other systems.

p. 20, line 4: It should be explained to what "improvements" compares to. Similar instruments by other groups use some of the "improved" conditions as well. Please be precise in your statements.

We added the references of the instruments to which we compared ours (page 23). We also added the minor difference of the second point with the study of Fuller et al. (2014): 2) separation of DCFH and peroxidase working solutions, which were then mixed just before reaction with the sample solution;

p. 20. Line 21: See above. The data presented here cannot support a lifetime estimate.

The reasons we would like to keep it are listed in the previous answer.

References

Berglund, G. I., Carlsson, G. H., Smith, A. T., Szoke, H., Henriksen, A., and Hajdu, J.: The catalytic pathway of horseradish peroxidase at high resolution, Nature, 417, 463-468, 2002.

Charrier, J. G. and Anastasio, C.: On dithiothreitol (DTT) as a measure of oxidative potential for ambient particles: evidence for the importance of soluble transition metals, Atmos. Chem. Phys., 12, 11317-11350, 2012.

Dall'Osto, M., Beddows, D. C. S., Harrison, R. M., and Onat, B.: Fine Iron Aerosols Are Internally Mixed with Nitrate in the Urban European Atmosphere, Environ. Sci. Technol., 50, 4212-4220, 2016.
Fuller, S. J., Wragg, F. P. H., Nutter, J., and Kalberer, M.: Comparison of on-line and off-line methods to quantify reactive oxygen species (ROS) in atmospheric aerosols, Atmos. Environ., 92, 97-103, 2014.

Janssen, N. A. H., Yang, A., Strak, M., Steenhof, M., Hellack, B., Gerlofs-Nijland, M. E., Kuhlbusch, T., Kelly, F., Harrison, R., Brunekreef, B., Hoek, G., and Cassee, F.: Oxidative potential of particulate matter collected at sites with different source characteristics, Sci. of The Total Environ., 472, 572-581, 2014.

Jiang, Y., Cui, C., Huang, Y., Zhang, X., and Gao, J.: Enzyme-based inverse opals: a facile and promising platform for fabrication of biocatalysts, Chem. Commun., 50, 5490-5493, 2014.

Kim, S., Jaques, P. A., Chang, M., Froines, J. R., and Sioutas, C.: Versatile aerosol concentration enrichment system (VACES) for simultaneous in vivo and in vitro evaluation of toxic effects of ultrafine, fine and coarse ambient particles Part I: Development and laboratory characterization, J. Aerosol Sci., 32, 1281-1297, 2001.

Oakes, M., Weber, R. J., Lai, B., Russell, A., and Ingall, E. D.: Characterization of iron speciation in urban and rural single particles using XANES spectroscopy and micro X-ray fluorescence measurements: investigating the relationship between speciation and fractional iron solubility, Atmos. Chem. Phys., 12, 745-756, 2012.

Perrone, M. G., Zhou, J., Malandrino, M., Sangiorgi, G., Rizzi, C., Ferrero, L., Dommen, J., and Bolzacchini, E.: PM chemical composition and oxidative potential of the soluble fraction of particles at two sites in the urban area of Milan, Northern Italy, Atmos. Environ., 128, 104-113, 2016.

Valko, M., Morris, H., and Cronin, M. T. D.: Metals, Toxicity and Oxidative Stress, Curr. Med. chem., 12, 1161-1208, 2005.

Zotter, P., Ciobanu, V. G., Zhang, Y. L., El-Haddad, I., Macchia, M., Daellenbach, K. R., Salazar, G. A., Huang, R. J., Wacker, L., Hueglin, C., Piazzalunga, A., Fermo, P., Schwikowski, M., Baltensperger, U., Szidat, S., and Prévôt, A. S. H.: Radiocarbon analysis of elemental and organic carbon in Switzerland during winter-smog episodes from 2008 to 2012 – Part 1: Source apportionment and spatial variability, Atmos. Chem. Phys., 14, 13551-13570, 2014.