Response to the comments of referee #1:

We thank the reviewer for the helpful comments.

Comment: In Table 4, the corrections for non-pseudo first-order conditions for the CRM instruments is given and appear to depend only on the overall reactivity. However, as discussed in Michoud et al. (2015), the correction also depends on the reactivity of the individual VOCs. It appears that the instruments always used an average correction factor for a range of reactive VOCs based on laboratory calibrations regardless of the VOC mixture in the chamber. This should be clarified. Could an inappropriate correction factor explain some of the discrepancies in the CRM measurements when specific VOCs were added to the chamber, such as the terpene mixture?

Response: The reviewer raises an important point: Indeed the corrections were made based on an average correction factor (typical of normal operating conditions), rather than based on the specific chamber mixture. The CRM systems have been developed for conditions prevalent in ambient air whereas in the chamber, artificial mixtures were used to test performance under simplified conditions (e.g. only CO or only monoterpenes). Instead of calibrating for the specific chemicals in the chamber, the CRM operators choose to apply their usual correction methods (average correction factor from the single component calibration bottles available to each group). This will indeed lead to a discrepancy particularly for situations where the reactivity of the injected compounds are significantly different from this average value as also discussed in the manuscript.

The correction for non-pseudo first-order conditions shown in Table 4 was calculated from the parameterization presented in Michoud et al. (2015) at a pyrrole-to-OH ratio of 2 for the MDOUAI instrument. This parameterization provides an average correction factor for trace gases exhibiting a bimolecular rate constant with OH ranging from 2.4×10^{-13} (ethane) to 10^{-10} (isoprene) cm^3 molecule^{-1} s^{-1}. The 1σ variability observed experimentally on this correction factor (when the different gases were used) is approximately 9%. As a consequence, the measured OH reactivity could deviate from the true reactivity by 9% (1σ) if a VOC (or group of VOCs) exhibiting OH rate constants close to the two bounds mentioned above were used in the chamber, with an overestimation and an underestimation of the OH reactivity for the less reactive and the most reactive compounds, respectively. The monoterpenes used in SAPHIR experiments exhibit rate constant ranging from 5.3×10^{-11} (α-pinene) to 2.1×10^{-10} (myrcene) cm^3 molecule^{-1} s^{-1}. Part of the discrepancy observed between the measurements of the MDOUAI instrument and the true OH reactivity in the chamber may therefore be due to an underestimation of the correction for non-pseudo first-order conditions. Since the rate constants of some of the monoterpenes used in SAPHIR are larger than the upper bound tested on the MDOUAI instrument, it is not straightforward to estimate the underestimation due to a miscorrection for non-pseudo first-order conditions. It would require testing compounds such as myrcene on the CRM instrument.

Comment: For the FT-LIF instrument, it was found that the measurements displayed greater deviations for more chemically complex mixtures, and the authors suggest that issues related to this particular flow tube instrument configuration may be responsible. What aspects of this particular instrument and its operation were significantly different from the normal PSU instrument?

Response: The Penn State group shipped their OH reactivity instrument to Juelich for the intercomparison, but had other obligations for the OH-measuring system, including the electronics, the drive controller for the OH reactivity instrument's movable wand, and the data collection system. Fortunately, Hartwig Harder's group at MPI Mainz was able to lend a laser system and to build a drive system. But this meant that new software had to be written to reduce the data. Also, the drive system had issues and needed to be reset often. The Penn State group also had to borrow a dye laser from Ron Cohen's group from UC Berkeley, which was similar to the one that Hartwig's group and we normally use, but not identical. In addition, the Penn State instrument normally pulls 100 LPM,

but it could not do that in the chamber. So they had to use a large dilution, which amplifies small systematic deviations and small day-to-day variations in the dilution flows by a factor of five or so.

Between the occasional problems with the laser and drive systems and the amplified variability due to the dilution, the instrument precision and the day-to-day accuracy was degraded somewhat from normal operation.

A more concise version of this response can already be found in the manuscript.

Comment: Two of the LP-LIF instruments observed bi-exponential behavior due to misalignment of the photolysis laser. How sensitive is the observed decay to the laser alignment? Is this sensitivity common to all LP-LIF instruments?

Response: The alignment of the photolysis laser in the flow tube is critical as shown in the results. Laboratory tests with the FZJ instruments, for example, have shown, that experimental parameters like laser beam diameter, beam profile, geometry of gas sampling by the LIF cell (head-on or sideways) have an influence on the decay curve. Once these geometrical parameters are properly fixed, the decay curves are usually reproducible over time and have a single exponential behavior. Deviations from this exponential behavior can be recognized, if the quality of the decay curve is checked, for example, by plotting on a logarithmic scale or by inspection of the residuum of the fit. On the other hand, reasonable corrections can be applied by post-processing as shown by the good agreement of all LP-LIF instruments in this campaign.

Response to the comments of referee #2:

We thank the reviewer for the helpful comments.

Comment: Both FZJ instruments should be briefly introduced in section 3.2. The FZJS instrument is used as the reference for the measurement with secondary reactions. It will be helpful to state the reason of the choice in more detail.

Response: All LP-LIF instruments are described together in Section 3.2. and differences between instruments are mentioned in the text. Subsections for the Lille and Leeds instruments are only describing the procedure how data were revised. Because this was not the case for the FZJ instrument, we do not think that a separate subsection for the FZJ instruments is needed.

The FZJS instrument was chosen as reference due to its good data coverage and high precision for experiments where it was not possible to calculate the OH reactivity. This was, for example, the case when the gas mixture in the chamber was exposed to solar radiation, and unmeasured products were formed. There are no other reasons than already described in the manuscript. Because the regression analysis takes errors in both coordinates into account, results are independent on the choice of the reference as stated on p17.

Comment: The accuracy and the precision of PSU LIF instrument used in the intercomparison was not as good as other LIF-based instrument. In line 18 on Page 15, the authors stated the reason was probably due to the dilution flow and the interference could be at least a factor of 5. More explanation will be expected, either in the paper or in the supplement materials.

Response: A similar question was raised in the first review and we would like to refer to our answer there. We think that there is no need for further explanation in the manuscript, because differences between the operation during this campaign and previous campaigns and the consequences for the measurement results are explained at various locations in the manuscript.

Comment: The sampling flow rates are quite different for all the instrument. The flow rates of DWD-CIMS and LP-LIF/FT-LIF are 2280 LPM and 10 to 20 LPM which the sample flow of CRM from SAPHIR chamber is not mentioned. Will it cause any interference on the measurement results?

Response: Sampling flow rate are listed in Table 1 for all instruments as well as the residence times in the inlet lines between chamber and instruments. As discussed for example on p23, the choice of material of the inlet line, residence time, temperature and surface to volume ratio of the inlet line potentially influence losses of OH reactants in the inlet line.

Comment: The summary ad conclusion gives a few nice points on various technology. It will be helpful if this sections can be separated into several parts.

Response: We will divide this part into a subsection "Summary of findings" and "Conclusions for future instrument operation and measurements in the past".

Response to the comments of referee #3:

We thank the reviewer for the helpful comments.

Comment: Authors comment that in previous field work large unexpected OH reactivity was found, both in biogenic and urban environments, but the results of the intercomparison showed that the contribution of OH reactivity of terpenoids and other oxygenated compound is understimated by CRM techniques. Could the authors extend the comments about this issue, and if the results of these intercomparison campaigns could lead to a re-interpretation of data results for some of the field works?

Response: An underestimation of the OH reactivity due to monoterpenes and OVOCs may have led to an underestimation of the missing OH reactivity when it was observed by CRM instruments during previous field campaigns. The reason for the underestimation is not known. It likely depends on the speciation and concentrations of the terpenoids and possibly experimental or ambient parameters. Without detailed knowledge, it is not possible to re-interpret data from these past campaigns since the monoterpenes and OVOCs mix was very likely different from that used during the present inter-comparison experiment. While this suggests that the reported missing reactivity represents a lower limit (especially in monoterpene rich environments), it seems to us that making extrapolations of the level of underestimation would be very hazardous.

Comment: VOCS from plant chamber: Could the authors give more experimental detail about experimental conditions e.g: flow used for plant emission transference to SAPHIR, humidity changes during experiment, number of trees used for the experiment, etc

Response: We will add:

- p7 l4: "Relative humidity typically dropped to 40 to 50 % during the experiment due to temperature changes and dilution."

- p7 l21: "In another experiment with biogenic reactants, emissions from plants (3 pine and 3 spruce, artificial light) were transferred into the chamber at a flow rate of $11\,m^3/h$."

- p7 l31: "...(typical range of relative humidity between 25 and 80 %)."

- p7 l32. "Initial ozone concentrations were around 100ppbv and dropped to 20ppb before re-injection."

Other minor comments will be included as suggested by the reviewer.

Comparison of OH reactivity measurements in the atmospheric simulation chamber SAPHIR

Hendrik Fuchs[1], Anna Novelli[1], Michael Rolletter[1], Andreas Hofzumahaus[1], Eva Y. Pfannerstill[2], Stephan Kessel[2], Achim Edtbauer[2], Jonathan Williams[2], Vincent Michoud[3,11], Sebastien Dusanter[3], Nadine Locoge[3], Nora Zannoni[4,a], Valerie Gros[4], Francois Truong[4], Roland Sarda-Esteve[4], Danny R. Cryer[5], Charlotte A. Brumby[5], Lisa K. Whalley[5,6], Daniel Stone[5], Paul W. Seakins[5,6], Dwayne E. Heard[5,6], Coralie Schoemaecker[7], Marion Blocquet[7,b], Sebastien Coudert[7], Sebastien Batut[7], Christa Fittschen[7], Alexander B. Thames[8], William H. Brune[8], Cheryl Ernest[2,c], Hartwig Harder[2], Jennifer B. A. Muller[9], Thomas Elste[9], Dagmar Kubistin[9], Stefanie Andres[1], Birger Bohn[1], Thorsten Hohaus[1], Frank Holland[1], Xin Li[1,d], Franz Rohrer[1], Astrid Kiendler-Scharr[1], Ralf Tillmann[1], Robert Wegener[1], Zhujun Yu[1], Qi Zou[10], and Andreas Wahner[1]

[1]Institute of Energy and Climate Research, IEK-8: Troposphere, Forschungszentrum Jülich GmbH, Jülich, Germany
[2]Air Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany
[3]IMT Lille Douai, Université Lille, SAGE - Département Sciences de l'Atmosphère et Génie de l'Environnement, Lille, France
[4]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France
[5]School of Chemistry, University of Leeds, LS2 9JT Leeds, UK
[6]National Centre for Atmospheric Science, University of Leeds, LS2 9JT Leeds, UK
[7]Lille, CNRS, UMR 8522 - PC2A - Physicochimie des Processus de Combustion et de l'Atmosphère, Lille, Université Lille, Villeneuve d'Ascq, France
[8]Department of Meteorology, Pennsylvania State University, University Park, PA, USA
[9]Meteorological Observatory Hohenpeissenberg, German Meteorological Service (DWD), Hohenpeissenberg, Germany
[10]College of Environmental Sciences and Engineering, Peking University, Beijing, China
[11]LISA, UMR CNRS 7583, Universités Paris Est Créteil et Paris Diderot, Institut Pierre Simon Laplace, Paris, France
[a]now at: Air Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany
[b]now at: Institute of Energy and Climate Research, IEK-8: Troposphere, Forschungszentrum Jülich GmbH, Jülich, Germany
[c]now at: University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany
[d]now at: College of Environmental Sciences and Engineering, Peking University, Beijing, China

Correspondence to: H. Fuchs (h.fuchs@fz-juelich.de)

Abstract. Hydroxyl (OH) radical reactivity (k_{OH}) has been measured for 18 years with different measurement techniques. In order to compare the performances of instruments deployed in the field, two campaigns were conducted performing experiments in the atmospheric simulation chamber SAPHIR at Forschungszentrum Jülich in October 2015 and April 2016. Chemical conditions were chosen either to be representative of the atmosphere or to test potential limitations of instruments. All types of instruments that are currently used for atmospheric measurements took part in one of the two campaigns. The results of these campaigns demonstrate that OH reactivity can be accurately measured for a wide range of atmospherically relevant chemical conditions (e.g. water vapor, nitrogen oxides, various organic compounds) by all instruments. The precision of the measurements (limit of detection $< 1\,\mathrm{s}^{-1}$ at a time resolution of 30 seconds to a few minutes) is higher for instruments directly detecting

hydroxyl radicals, whereas the indirect Comparative Reactivity Method (CRM) has a higher limit of detection of $2\,\mathrm{s}^{-1}$ at a time resolution of 10 to 15 min. The performances of the instruments were systematically tested by stepwise increasing, for example, the concentrations of carbon monoxide (CO), water vapor or nitric oxide (NO). In further experiments, mixtures of organic reactants were injected in the chamber to simulate urban and forested environments. Overall, the results show that

5 instruments are capable of measuring OH reactivity in the presence of CO, alkanes, alkenes and aromatic compounds. The transmission efficiency in Teflon inlet lines could have introduced systematic errors in measurements for low-volatile organic compounds in some instruments. CRM instruments exhibited a larger scatter in the data compared to the other instruments. The largest differences to reference measurements or to calculated reactivity were observed by CRM instruments in the presence of terpenes and oxygenated organic compounds (mixing ratio of OH reactants were up to 10 ppbv). In some of these

10 experiments, only a small fraction of the reactivity is detected. The accuracy of CRM measurements is most likely limited by the corrections that need to be applied in order to account for known effects of, for example, deviations from pseudo-first order conditions, nitrogen oxides or water vapor on the measurement. Methods to derive these corrections vary among the different CRM instruments. Measurements by a flow-tube instrument combined with the direct detection of OH by chemical ionization mass spectrometry (CIMS) show limitations in cases of high reactivity and high NO concentrations, but were accurate for low

15 reactivity ($< 15\,\mathrm{s}^{-1}$) and low NO (< 5 ppbv) conditions.

1 Introduction

Most gas species in the atmosphere are transformed by their reaction with the hydroxyl radical (OH). These processes lead to the formation of oxidized, secondary pollutants such as ozone and aerosol. Due to the large number of organic OH reactants (Goldstein and Galbally, 2007), several methods have been developed in order to measure OH reactivity (the inverse OH lifetime). OH reactivity (k_{OH}) is the sum of OH reactant concentrations ([X]) weighted by their reaction rate coefficient with OH ($k_{\mathrm{OH}+X}$):

$$k_{\mathrm{OH}} = \sum_i k_{\mathrm{OH}+Xi}[\mathrm{X_i}] \tag{1}$$

Predicting trace gas loadings and lifetimes requires a comprehensive understanding of the atmosphere's chemical cycling and oxidative capacity, which is aided by the measurement of total OH reactivity. Measurements can be compared to calculations from OH reactant concentrations in order to quantify unexplained reactivity. In addition, the total loss rate of OH can be calculated, if OH concentrations are concurrently measured in order to analyze the OH budget by comparing the total OH loss rate with the sum of OH production rates.

The measurement of OH reactivity has been shown to be extremely useful (Yang et al., 2016). Up to several ten per second unexplained reactivity was, for example, identified in biogenic dominated environments such as in a forest in Michigan (Di Carlo et al., 2004; Hansen et al., 2014), in the Amazonian rain forest (Nölscher et al., 2016) and in a boreal forest in Finland (Nölscher et al., 2012b). The magnitude of missing reactivity appears to be dependent on the biogenic source, time of the day and season (Williams et al., 2016; Nölscher et al., 2016). However, the agreement between measured and calculated reactivity is also a valuable result because it indicates that all trace gases that are relevant for the photochemistry were measured. This was the case in environments that were influenced by anthropogenic OH reactants as in New York (Ren et al., 2003) and in the North China Plain (Fuchs et al., 2016), in isoprene-dominated environments during daytime in a Mediterranean forest (Zannoni et al., 2016), and in a chamber study (Nölscher et al., 2014). In addition, the gap between measured and calculated OH reactivity could be closed in some field studies, if oxygenated VOCS derived from model calculation were additionally taken into account (e.g. Chatani et al., 2009; Lou et al., 2010; Kaiser et al., 2016; Whalley et al., 2016). First attempts were also made to measure OH reactivity fluxes (Nölscher et al., 2013).

The application of OH reactivity measurements for the analysis of the OH budget also provided new results. A gap in the understanding of OH recycling processes was found in a field study in Nashville in 1999 (Martinez et al., 2003), in China in 2006 (Hofzumahaus et al., 2009), in Borneo in 2008 (Whalley et al., 2011) and in chamber experiments investigating the oxidation of isoprene by OH (Fuchs et al., 2013). Because of the close connection between oxidation of organic compounds by OH and ozone production, OH reactivity can help to calculate local ozone production rates (Sinha et al., 2012).

Several methods to measure OH reactivity have been developed since the first measurements were made by Penn State University (PSU) (Kovacs and Brune, 2001). The different methods fall into two categories. One method determines the OH reactivity directly from the time-dependent decay of measured OH that is artificially produced. The other method determines

k_{OH} indirectly from the concentration change of a reference species which competes with atmospheric reactants in their reaction with artificially produced OH.

In the instrument developed by Kovacs and Brune (2001), the decay of OH is measured in a flow tube through which ambient air is drawn by the direct detection of OH using laser-induced fluorescence. OH is continuously produced by water photolysis.

5 The time-resolved OH decay is measured by varying the reaction time using a movable injector to produce OH. A compact aircraft instrument was later developed by PSU and deployed for the first time in 2006 (Mao et al., 2009). Similar instruments were built at Indiana University (Hansen et al., 2014) and at University of Leeds (Ingham et al., 2009). The latter apparatus was recently replaced by an instrument applying a pump-probe technique (see below).

In an alternative instrument a flow tube set up is combined with a chemical ionization mass spectrometer (CIMS) which

10 detects sulfuric acid (H_2SO_4) following the chemical conversion of OH to H_2SO_4 (Berresheim et al., 2000; Muller et al., 2017). In this instrument developed by the German Meteorological Service (DWD), OH is produced by water photolysis in the flow tube. The reaction of OH with ambient OH reactants is terminated by chemically removing OH by its reaction with sulfur dioxide that is injected at two positions within the flow tube giving one reaction time for the OH decay. The remaining OH concentrations for the two injection positions are measured to calculate the OH reactivity.

15 Sadanaga et al. (2004) developed an instrument that uses a pump-probe technique, called Laser Photolysis - Laser Induced Fluorescence (LP-LIF). OH is produced by ozone photolysis using radiation of short laser pulse at $266\,nm$ at a low repetition rate of 1 to $2\,Hz$. The OH decay is observed by laser-induced fluorescence with a high time resolution. The pump-probe technique has the advantage that the flow conditions do not need to be exactly known in order to determine a reaction time. This technique is now used by several groups such as Tokyo Metropolitan University (Sadanaga et al., 2004), University of

20 Leeds (Stone et al., 2016), University of Lille (Parker et al., 2011), and Forschungszentrum Jülich (FZJ) (Lou et al., 2010).

The indirect technique for the measurement of OH reactivity was pioneered by Sinha et al. (2008). The Comparative Reactivity Method (CRM) is based on the detection of pyrrole that reacts with artificially produced OH in clean or ambient air. The pyrrole competes with the ambient OH reactants, so that the pyrrole concentration depends on the ambient OH reactivity. In most CRM instruments, pyrrole is detected by proton-transfer-reaction mass-spectrometry (PTR-MS), but can also be detected

25 by gas-chromatography (GC) (Nölscher et al., 2012a). The CRM method is more commonly used than the direct OH measurement techniques because of the commercial availability of PTR-MS instruments. It is applied by Max-Planck-Institute Mainz (MPI) (Sinha et al., 2008), IMT Lille Douai, formally called Mines Douai, (MDOUAI) (Hansen et al., 2015; Michoud et al., 2015), Laboratoire des Sciences du Climat et de l'Environnement (LSCE) (Zannoni et al., 2015), Indian Institute of Science Education and Research Mohali (Kumar et al., 2014), Finish Meteorological Institute (Praplan et al., 2017), Peking University

30 (Yang et al., 2017), University Leicester and University of California, Irvine (Kim et al., 2016).

Only two side-by-side comparisons have been performed between two CRM instruments in a remote environment (Zannoni et al., 2015) and between a CRM and a pump-probe instrument in an urban environment (Hansen et al., 2015). Both comparisons show generally good agreement between measurements within 20 to $50\,\%$.

In 2014 a workshop was held at Max-Planck Institute for Chemistry in Mainz in order to assess the current status of and

35 future for OH reactivity measurements (Williams and Brune, 2015). At the workshop, a comparison campaign was suggested

to investigate the performance of instruments under different atmospheric chemical conditions. Large environmental chambers are ideal for this purpose as they ensure that all instruments sample air with the same chemical composition. In addition, chemical conditions can be systematically varied. This was demonstrated in several comparison exercises in the atmospheric simulation chamber SAPHIR at Forschungszentrum Jülich (e.g. Schlosser et al., 2009; Dorn et al., 2013) as well as in the EUPHORE chamber (e.g. Pang et al., 2014). Here, we report the results of two k_{OH} comparison campaigns that were conducted in the SAPHIR chamber. The two comparisons were not blind, quick-look data were presented from some groups during the campaigns. After first data submission without the knowledge of final results from other participants or OH reactant concentrations, data were allowed to be revised. Only final data are presented in this paper, but changes after the first data submission are described.

A large number of OH reactivity instruments applying different techniques successfully participated in these campaigns (CRM instruments of MPI, IMT Lille Douai and LSCE; a flow tube LIF instrument of PSU; a CIMS instrument from DWD and LP-LIF instruments of Lille, Leeds and FZJ, Table 1). The CRM instrument of Finish Meteorological Institute also took part in the campaign, but measurements by this instrument failed due to technical problems and no valid data could be acquired.

2 Experiments in the SAPHIR chamber

2.1 The SAPHIR chamber

The outdoor atmospheric simulation chamber SAPHIR is made of a double-wall Teflon (FEP) film and has a cylindrical shape (5 m diameter, 18 m length). The Teflon chamber is mounted inside a steel frame that is equipped with a shutter system that allows for experiments in the dark or in the presence of sunlight. The space between the inner and outer Teflon film is continuously purged with nitrogen (Linde, purity > 99.9999 %) to prevent contamination from outside. In addition, the pressure inside the chamber is 45 Pa higher than ambient pressure. Small leakages and air sampling by instruments require replenishing the air to maintain the pressure difference. This leads to a small dilution of trace gases by 3 to 5 % per hour. The dilution can be as high as 60 % per hour if the replenishment flow needs to be high.

Ultra-pure air (Linde, purity > 99.9999 %) is used to purge the chamber with a high flow (up to 250 m^3/h) in order to clean the chamber. The high flow rate is also required to humidify the chamber air with steam from boiling water that is supplied by a Milli-Q water device. Ozone produced by a silent discharge ozonizer can be added to the chamber air. Two fans ensure that the air is well mixed, so that all instruments always sample the same air mass (e.g. Schlosser et al., 2009; Dorn et al., 2013).

The use of high purity air ensures that there are no measurable gaseous OH reactants present in the chamber after the purging procedure. Small amounts of mostly unidentified organic compounds and also nitrogen oxide compounds like HONO (< 100 pptv) can be observed in some cases during the humidification. The total OH reactivity measured by the OH reactivity instrument that is permanently installed at the chamber shows that the reactivity is typically well below 1 s^{-1} after humidification. In fact, instruments measured on average no significant OH reactivity in these campaigns in the clean chamber (see below).

If the chamber is exposed to sunlight, well-characterized sources for HONO, formaldehyde and acetaldehyde lead to the increase of OH reactivity (production rates are typically $200\,\mathrm{pptv/h}$). Photolysis of HONO (Rohrer et al., 2005) is also the dominant source of OH and nitrogen oxides in the chamber. The source strength depends on the relative humidity, temperature and radiation. The overall increase of OH reactivity is on the order of $0.2\,\mathrm{s}^{-1}$ per hour, and is much smaller than the reactivity

5 from added OH reactants in these campaigns.

OH reactants were added either from gas mixtures via calibrated flow controllers or as liquids that were injected into a heated inlet line with a syringe. The vapors were transported by a flow of synthetic air into the chamber. In addition, a recently built plant chamber allows for the quantitative transfer of mixtures of biogenic organic compounds from up to six trees into the SAPHIR chamber (Hohaus et al., 2016). Environmental parameters in the plant chamber can be fully controlled.

10 **2.2   Calculated OH reactivity**

A number of instruments for the detection of OH reactants measured concurrently with the OH reactivity instruments (Table 2). Nitrogen oxides (NO and NO_2) were detected by a chemiluminescence instrument (Eco Physics TR 780). CO was measured by a Piccarro cavity ring-down instrument (Picarro G2301) and also by gas-chromatography (GC, Trace Analytical RGA 3). Both measurements agreed within $5\,\%$. Data from the cavity ring-down instrument were used here for calculations of

15 the OH reactivity due to its higher accuracy. This instrument also measured methane and water vapor concentrations. Organic compounds were detected by PTR-TOF-MS (proton-transfer-reaction time-of-flight mass-spectrometry, Ionicon) and GC (Agilent 7890N). Measurements agreed for those species that can be detected by both instruments such as isoprene better than $20\,\%$ with some larger discrepancies for acetaldehyde and β-caryophyllene in the second set of experiments in 2016 (Table 2). Differences between measurements need to be regarded as additional uncertainty in the calculation of OH reactivity.

20   PTR-TOF-MS measures the sum of methyl vinyl ketone (MVK) and methacrolein (MACR) and the sum of monoterpenes. In order to calculate OH reactivity, PTR-TOF-MS measurements were used taking the relative distribution of MVK and MACR and monoterpenes as measured by GC, because PTR-TOF-MS has a high time resolution. Formaldehyde was additionally measured by a Hantzsch monitor (Aero Laser AL 4001). All reaction rate constants used for the calculation of OH reactivity are taken from IUPAC recommendations (Atkinson et al., 2004, 2006; IUPAC, 2017), if not stated differently in Table 2. Tem-

25 perature and pressure are assumed to be the same in the instruments and the SAPHIR chamber. This approach is applicable as indicated by temperature and pressure measurements in the instruments. The overall 1σ uncertainty of the calculated OH reactivity is around $20\,\%$ in most experiments, but can be higher (e.g. $40\,\%$ in case of the experiment with sesquiterpenes) depending on the uncertainty in the OH reactant measurements, the agreement between simultaneous measurements by different instruments and the uncertainty in reaction rate constants.

30 **2.3   Experiments performed in 2015**

Two campaigns were conducted for this comparison. The first one took place in October 2015. All instruments listed in Table 1 took part in this campaign with exception of the CIMS instrument.

In the experiments, the chamber was flushed with high purity air before each experiment, until trace gas concentrations were below the limit of detection of instruments (Table 2). The chamber air was humidified to approximately 75 % relative humidity (RH) at the beginning of each experiment, except for the experiment on 06 October 2015, when the experiment was started with 25 % RH and the humidity was increased to 90 % RH in three steps. Relative humidity typically dropped to 40 to 50 %

5   during the experiment due to temperature changes and dilution. Ozone was also added at the beginning of the experiments to allow OH production in the LP-LIF instruments if ozone was not expected to affect the chemical composition of the chamber air (e.g. by ozonolysis reaction or by the conversion of NO to NO_2). Initial ozone mixing ratios were typically between 50 and 80 ppbv.

OH reactivity was typically increased in several steps to maximum values of approximately $50\,s^{-1}$ at the end of the exper-

10   iment (maximum $150\,s^{-1}$). The time between two injections of trace gases was approximately 45 min. In addition, chemical conditions were changed during the course of some experiments, such as opening or closing the chamber roof or adding nitrogen oxides or water vapor. Chemical conditions in the different experiments are summarized in Table 3.

Some experiments aimed primarily for testing the instruments' performances: linearity with CO (05 October 2015), the influence of humidity (06 October 2015) and the presence of NO (07 October 2015) were tested. The last test was repeated

15   on 15 October. However, due to an operational error, ozone was added at the beginning of the experiment, so that a mixture of NO and NO_2 was present. In order to reduce the ozone concentration, 2,3-dimethyl-2-butene (TME), which reacts rapidly with ozone, was injected twice.

The other experiments focused on the instruments' performances in the presence of specific OH reactants and atmospheric mixtures of reactants. In part of these experiments, OH reactants were also oxidized by either OH or ozone. In five experiments,

20   biogenic reactants were present: isoprene (11 October 2015), isoprene oxidation products MVK and MACR (16 October 2015), a mix of monoterpenes (α-pinene, limonene, myrcene, 09 October 2015), and a sesquiterpene (β-caryophyllene, 16 October 2015). In another experiment with biogenic reactants, emissions from plants (3 pine and 3 spruce)~~(pine and spruce)~~ were transferred into the chamber at a flow rate of $11\,m^3/h$. In two experiments, an urban environment was simulated by a mixture of 1-pentene, o-xylene and toluene together with NO_2 (12 and 13 October 2015). On 16 October 2015, acetaldehyde was

25   injected.

2.4   Experiments performed in 2016

In the second campaign in 2016, only three instruments measured OH reactivity: a CRM instrument (MPI), a LP-LIF instrument (FZJS) and the CIMS instrument of DWD. The CRM and LP-LIF instruments were the same as in the 2015 campaign. The CIMS instrument sampled air with a high flow rate (2280 liter/min) requiring the chamber to be operated with a high

30   replenishment flow. As a consequence, all trace gases were diluted at a high rate of approximately 60 % per hour. Oxidation products could not accumulate. Accordingly, the experimental procedure was different in these experiments compared to those in 2015: Humidification was done two to four times over the course of an experiment in order to maintain a sufficiently high water vapor concentration for the production of OH in the LP-LIF and CIMS instruments (typical range of relative humidity between 25 and 80 %). If ozone was present in the experiment, ozone was also injected several times. Initial ozone concentra-

tions were around 100 ppbv and dropped to 20 ppbv before re-injection. Similar chemical conditions as in the first campaign were tested in order to achieve comparable results. Tests were done with single, anthropogenic OH reactants (CO, pentane) in the presence (08 April 2016) and in the absence (07 and 15 April 2016) of NO, with biogenic reactants (isoprene, MVK, MACR on 09 April 2016, a monoterpene mixture on 13 and 14 April 2106, β-caryophyllene on 15 April 2016) and with a mixture of anthropogenic reactants (11 and 12 April 2016). The same monoterpene and urban reactant mixtures were used as in the experiments in 2015 and in 2016.

2.5 Data coverage

The CRM instrument of the Finish Meteorological Institute (FMI) took part in the first campaign, but no valid measurements could be acquired due to technical problems of this instrument. Data were submitted for all other instruments for nearly all experiments and are included in the comparison. A leak in the OH injection system of the MDOUAI CRM instrument was found after the third experiments and this leak possibly led to systematic errors in the measurements in this experiment. Data from the experiment on 07 October 2015 were therefore rejected for this instrument. The sampling system of the CRM instrument from LSCE was changed on the second day of the campaign (06 October 2015). Measurements from this experiment were rejected. On 12 October 2015, the flow-tube instrument from PSU did not measure except for the last two hours due to technical problems. All other instruments measured during all times of the campaign.

2.6 Procedure of data comparison

The measurement comparison was not strictly blind, but some rules were applied to which all participants had agreed prior to the campaign. The general outline of the campaign was:

- Before the official campaign started, a test experiment with CO was performed in the SAPHIR chamber. During this experiment, the participants were informed about the added CO concentrations in order to test the functionality of their instruments (04 October 2015, not included in the comparison, and 07 April 2016).

- During the campaign, participants were informed about the type of trace gases which were planned to be added to the chamber air before the experiments. Concentrations of reactants, however, were not disclosed to the participants.

- During the campaign, participants had the opportunity to present quick-look data of measured values at daily meetings, but data were not exchanged or distributed.

- After the campaign, all participants independently submitted their evaluated data to a neutral person at Forschungszentrum Jülich who was not involved in reactivity measurements. Only after all data were received, the measured trace gas concentrations and the k_{OH} data of all participants were made available.

- After data disclosure, some participants applied corrections to their data and submitted a revised data set together with an explanation for the correction.

- The comparison in this paper is based on the final data versions.

Changes of data that were done as a result of the comparison are described in the next section for each instrument.

3 Instruments for the detection of OH reactivity

3.1 Comparative Reactivity Method (CRM)

The Comparative Reactivity Method (CRM) is an indirect method for the measurement of OH reactivity developed by Sinha et al. (2008). The measurement principle relies on the competition of the reaction of OH with either a known pyrrole concentration or ambient OH reactants. Pyrrole acts as a reference species that is typically not present in ambient air. A small flow of humidified, ultra-pure nitrogen (flow rate approximately $240\,\mathrm{cm^3/min}$) is flowed over a pen-ray lamp leading to formation of OH by water photolysis at $185\,\mathrm{nm}$ with concentrations of approximately 1 to $3 \times 10^{12}\,\mathrm{cm^{-3}}$ (Table 1). Water photolysis, however, not only produces OH, but also HO_2 radicals. The higher reactivity of OH compared to HO_2 also towards surfaces may lead to HO_2 concentrations exceeding the concentration of OH in the reactor.

Ambient OH reactants and/or pyrrole react with OH in a reaction volume ($94\,\mathrm{cm^3}$) made of glass with the inner surface covered by Teflon. The instrument is alternately switched between two measurement modes: the small flow of pyrrole (approx. 2 to $3\,\mathrm{cm^3/min}$) is mixed into a flow of purified air (C2-mode) (approximately $300\,\mathrm{cm^3/min}$) or into a flow of ambient air (air sampled from the chamber in these experiments) (C3-mode). As OH exclusively reacts with pyrrole in the C2-mode, maximum reduction of the pyrrole concentration is achieved, whereas the pyrrole concentration is higher in the C3-mode, when also ambient OH reactants are present. In order to calculate the OH reactivity, the initial pyrrole concentration needs to be known (typically 1 to $2 \times 10^{12}\,\mathrm{cm^{-3}}$). Because a small fraction of the radiation of the pen-ray lamp enters the reaction volume, a small fraction of the pyrrole is photolyzed (typically less than $10\,\%$). Therefore, the pyrrole concentration is measured, when zero air is sampled and when the light of the pen-ray lamp is turned on (C1-mode). This is typically done once a day.

The design of the reaction volume is identical for all instruments, because they were all manufactured by the Max-Planck-Institute for Chemistry in Mainz. Three CRM instruments are included in this comparison: The instruments of MPI, IMT Lille Douai (MDOUAI) and LSCE. The instruments differ mainly in the exact operational conditions such as flow rates, pyrrole and OH concentrations and the inlet lines (Table 1). The transformation of raw data into k_{OH} values requires corrections (Michoud et al., 2015) that have been characterized for each CRM instrument (Table 4). These corrections, described below, can significantly differ between instruments due to the different operating conditions.

The pyrrole concentration is monitored by Proton-Transfer-Reaction Mass-Spectrometry (PTR-MS) in nearly all instruments, but can also be detected by gas-chromatography (GC) (Nölscher et al., 2012a) . This is for example done in the instrument of Finish Meteorological Institute.

A number of corrections need to be applied to the signals measured in the different modes due to a variety of factors (Michoud et al., 2015):

- The OH production rate in the two measurement modes can be different, if the water vapor concentration is not the same in both modes.

- OH can be significantly reformed by the reaction of HO_2 that is present at high concentrations in the reactor with ambient NO.

- The reaction deviates from pseudo-first order conditions.

- Ambient OH reactant concentrations are diluted due to the additional nitrogen flow. The dilution factor is calculated from measured flow rates.

Corrections are usually determined from instrumental characterization in the laboratory and in the field, with the assumption that they are representative for ambient air measurements. Typical values of corrections are listed in Table 4.

10 All groups operating a CRM used empirical functions to correct for deviations from the pseudo-first-order decay for the final data evaluation (Michoud et al., 2015). The exact value, however, depends on the chemical composition of OH reactants (see below). Different representative mixtures are taken for characterization of this correction for the various CRM instruments and operating conditions are optimized to reduce the correction dependence on the chemical composition. The error associated to this correction can then be factored in the measurement uncertainty (Michoud et al., 2015).

15 Additional instrument specific corrections are described in the following section.

3.1.1 MPI CRM instrument

The correction of measurements of the CRM by MPI for deviations from a pseudo-first reaction used to done by using results from numerical simulations. However, this can only be applied if the relative importance of the most abundant reactive compounds in the sampled air is known (Sinha et al., 2008). The model correction was not applicable in this comparison because

20 no data on the concentration ratios of the main OH reactants were known in contrast to typical situations in field campaigns. In this campaign, also the empirical correction procedure was chosen as an alternative correction procedure that was shown to be advantageous by Michoud et al. (2015). Tests with isoprene, methanol, ethane, propane, propene and toluene were done to determine the correction factor.

In addition to the corrections applied by all groups operating a CRM instrument, measurements by the MPI CRM were

25 corrected for the presence of ambient ozone. The necessity for this correction was recognized after the first comparison of results from the 2015 campaign. This correction was not applied in the first version of submitted data. The procedure to correct data were then determined in laboratory characterization experiments. The correction was applied to data of the 2016 campaign from the beginning.

OH is reformed in the reaction of HO_2 with O_3 in the reaction volume of the CRM, where O_3 is present in sampled ambient

30 air, but is also produced in the photolysis of oxygen by the 185 nm radiation of the pen-ray lamp. The assumption is that the effect of OH reformation on the measurement is typically insensitive to the exact concentration of ambient ozone, which is not present in all modes of the measurement cycle. If this assumption is not true, the OH reactivity is underestimated depending on

the ambient ozone concentration. This was observed for the MPI CRM instrument in this campaign. Therefore, measurements were corrected by an empirical function derived from laboratory measurements after the first data submission. Although the ozone concentration in the reaction volume was similar to the concentration in the other two CRM instruments, no ozone dependence was seen for the MDOUAI and LSCE instruments. The exact reason is not clear, but might be related to different HO_2 concentrations in the instruments. The insensitivity of other CRM instruments to the ozone interference indicates that operating conditions exist for which the interference is negligible. Further investigations should be performed to characterize these conditions.

In addition to the ozone correction, errors in the calculation of the dilution factor and the calibration of the pyrrole sensitivity were noticed for the MPI CRM instrument after the first data submission in 2015. Although corrections were done after knowledge of OH reactant concentrations and measurements of other instruments, unreasonable data already suggested the need for these corrections before. Also the correction for the presence of NO_2 was again characterized for conditions when also O_3 was present and slightly changed in the final data. Furthermore, the correction due to the deviations from pseudo-first order decay were changed in the final data because a re-analysis of the concentration of test compounds used for the characterization revealed higher impurities than certified by the manufacturer.

The total increase of OH reactivity measurements between the first and final submission was typical within the range of a factor of 1.5 to 3, but was a factor of 4 to 5 for low OH reactivities around $10\,\mathrm{s}^{-1}$ in the presence of ozone mixing ratios of $40\,\mathrm{ppbv}$.

3.1.2 MDOUAI CRM instrument

The performance of the MDOUAI CRM instrument was worse in this campaign than previously observed due to additional sources of noise from the PTR-MS instrument and the inlet system. It was recognized that the pump (Teflon surfaces) in the sampling line upstream of the CRM instrument, which is necessary to avoid a pressure drop between ambient pressure and the CRM reactor (Michoud et al., 2015), released contaminants which caused an additional OH reactivity of $15\,\mathrm{s}^{-1}$ on average during measurements. This instrument zero was subtracted from all measurements. The value was determined daily in each experiment between humidification of chamber air and injection of OH reactants. Deviations from pseudo-first order behavior of the kinetics were characterized by tests with isoprene, propene, ethene, ethane and propane. Data were not revised after the first data submission.

3.1.3 LSCE CRM instrument

At the beginning of the campaign, a pressure change was observed for the two measurement modes of the CRM instrument at the exit of the reactor that could affect the measurements. The total flow rate in the sampling line was increased and only a small part was sampled into the reaction volume in order to avoid a change in pressure. Therefore, the sampling flow was not directly injected into the CRM reactor but it was first pulled through a pump with Teflon surfaces. The flow was restricted by a valve (Teflon surfaces) before the air entered the reactor. This sampling flow system was used for the first time during this campaign and may have reduced the performance of the instrument.

Corrections were applied to the raw data as described by Michoud et al. (2015) to obtain the OH reactivity values . Specifically, the correction for the deviation from the pseudo-first order conditions was determined from laboratory and field tests using certified concentration of gas standards containing propane and isoprene. The same procedure was applied in a previous field campaign in an isoprene dominated environment (Zannoni et al., 2016). Previous field deployments of the same instrument were conducted in environments with low NO_x concentrations, for which a correction for OH recycling by NO was not needed. For this reason, a correction for high NO_x concentrations was determined in laboratory tests after the campaign in SAPHIR and data of the experiments from LSCE were revised after the first submission for experiments, when NO_x was injected.

Instrument operators decided to use the part of the experiments before OH reactants were added to subtract a background, when positive, non-zero values were on average measured (05, 07, 09, 16 October 2015), in their first data evaluation. However, this correction was not applied in the final data set, because it was agreed not to use knowledge of the chemical conditions for the data evaluation, if it is not required.

Changes in the revised OH reactivity measurements were smaller than $\pm 20\%$ except for measurements at high NO mixing ratios on 07 October 2015, when changes were up to $80\,\mathrm{s}^{-1}$ as no correction for the NO interference was applied in the first submitted data.

3.2 Direct OH loss rate measurement by Laser-photolysis laser-induced fluorescence (LP-LIF)

All other instruments measured the decay of OH in the presence of ambient OH reactants in a flow tube. In most of the instruments, OH radicals are detected by LIF (Heard and Pilling, 2003).

All methods measuring the OH decay have a higher time resolution compared to the CRM instruments (Table 1) because no time is consumed switching between ambient and purified air. In general, fewer corrections are required to derive the OH reactivity from the measured OH decay.

Four LP-LIF instruments took part in the campaigns: the instruments of University of Lille and University of Leeds and two instruments of FZJ, one of which is permanently installed at the SAPHIR chamber (FZJS) and the other one is used for mobile field deployment (FZJM).

In the laser-photolysis instruments, ambient air is flowed (flow rate 10 to $20\,\mathrm{l/min}$) through a flow tube. Part or all of the air is drawn into an OH fluorescence detection cell. The exact position and design of the flow tube and the fluorescence cell differs among the instruments. OH is produced by flash photolysis of ozone with subsequent reaction of $O(^1D)$ with water vapor. Radiation is provided by a quadrupled Nd:YAG laser pulse at $266\,\mathrm{nm}$, which is operated at a low repetition rate of 1 to $2\,\mathrm{Hz}$. The OH concentration is measured with a high frequency of 3 to $8.5\,\mathrm{kHz}$, so that the decay of OH can be observed with a high time resolution between two photolysis shots. Tens of consecutive decays are summed up to increase the signal to noise ratio.

The OH radicals decay in a pseudo-first order reaction with ambient OH reactants, so that the time resolved OH signal can be fitted to a single exponential function that directly gives the OH reactivity. Differences between the fitting procedures of the

instruments are described in the supplementary information. The accuracy of the time basis of the OH decay is only determined by the accuracy of the photon counting electronics.

Measurements need to be corrected for an instrument zero that is subtracted from all measurements. This zero loss rate in the flow tube is due to the wall loss reactions and likely limited by the diffusion of OH. Values are typically on the order of a

5   few s^{-1} (Table 4) and are regularly determined by sampling high purity zero air.

Conversion of HO_2 to OH in the presence of ambient NO can influence the measurements. As there is no concurrent production of HO_2 in the ozone photolysis, LP-LIF instruments are less affected by this recycling process compared to instruments that use water photolysis for OH production. It is expected that this recycling process only becomes relevant for NO mixing ratios higher than 20 ppbv for typical atmospheric chemical compositions of ambient air (Lou et al., 2010). In this case, the

10   single exponential decay of OH turns into a bi-exponential decay that can clearly be identified in the summed decays. If a bi-exponential decay is observed, the faster decay time can be attributed to the OH reactivity. The underlying assumption is that the time scale of OH formation is slow compared to the OH loss. This is reasonable for typical atmospheric conditions, but may not be applicable in all cases, specifically in artificial air mixtures. In field experiments, bi-exponential behavior in the OH decay due to OH recycling at high NO concentrations in ambient air measurements has been observed by the FZJS and

15   Lille instruments. A bi-exponential function was applied for measurements in a campaign in China for the FZJS instrument (Lou et al., 2010). Measurements of the Lille instrument that showed a bi-exponential behavior during a campaign on the campus of the University Lille were evaluated by applying a single exponential function. Measurements were evaluated by only using the first part of the decay curve that contained the information of the faster decay (Hansen et al., 2015). No significant difference between this procedure and results from a bi-exponential fit was found. The fitting of the data using a single or a

20   bi-exponentially decay function is discussed later in the manuscript, as differences were observed in the returned value of the OH reactivity in this campaign at high NO_x (> 20 ppbv) depending on the type of fit used.

In the normal operational procedure, zero or only a small dilution needs to be taken into account for most of the instruments. If there is insufficient ambient ozone to generate a measurable OH signal, then an addition of O_3 containing flow needs to be added to the flow tube and a small dilution correction needs to be made. This was required in experiments without the presence

25   of ozone in the chamber air.

The number of data points on the decay curve that are above the noise level decreases with increasing OH reactivity, so that accuracy and precision of measurements start to decrease for exceptionally high OH reactivity values (for example higher than $60\,s^{-1}$ for the FZJ LP-LIF instrument). In addition, initial inhomogeneities in the OH distribution in the flow tube due to inhomogeneities of the laser photolysis beam can impact the shape of the observed OH decay curve for these high reactivity

30   values. For this reason, an additional dilution flow can be applied in order to reduce the OH reactant concentrations and improve the data quality. This was done in some instruments (FZJ, Lille) in this campaign, when the measured reactivity exceeded a threshold (e.g. $> 150\,s^{-1}$ for the Lille instrument), but is not required as indicated by measurements by the Leeds LP-LIF instrument.

Imperfect alignment of the photolysis laser can enhance the inhomogeneities in the initial OH distribution, so that deviations

35   from a single exponential OH decay can also occur at low reactivity values. This was observed in this campaign in the Lille

and Leeds LP-LIF instruments, but recognized only at the end of (Lille) or after (Leeds) the campaign. As a consequence, the evaluation procedures were changed for measurements in this campaign in order to account for this effect.

Data of FZJS and FZJM instruments were not revised after the first data submission and no instrument specific description is required here. The Lille and Leeds instruments required a campaign specific data evaluation that was applied before (Lille) or after (Leeds) the first data submission.

3.2.1 Lille LP-LIF instrument

Quick-look data presented from the LP-LIF instrument of Lille systematically deviated from measurements of the other instruments. The overestimation of approximately 30 % was confirmed by determining the reaction rate constant of the reaction between CO and OH in test experiments, in which a mixture of CO in synthetic air was sampled. This overestimation was due to two reasons: (1) misalignment of the photolysis laser leading to deviations from a single-exponential behavior of the OH decay curve likely due to an inhomogeneous initial OH concentration (see above); (2) the procedure of analyzing the decay by adapting the length of the decay curve used for the fit. The length is limited to 15 times of a first estimate of the inverse reactivity (Hansen et al., 2015) in order to avoid noise from the background signal at longer times. As a consequence, the fitted zero decay time appeared to change, if the length of the curve used for the fit was shortened for zero air measurements like done for the high reactivity values. This was then utilized to account for the deviations in the reactivity measurements by determining an artificial zero decay time as a function of the fit length. In the final data, this zero decay time which depends on the fit length and therefore reactivity value was subtracted from the measurements (Fig. 1 in the supplementary information). With this method, correct reactivity values could be calculated for the laboratory test experiments with CO. The drawback is that the accuracy is less for high reactivity values due to the decreasing number of points used for the fit.

This correction would not be needed for good alignment of the photolysis laser. It is therefore only needed for the data evaluation of this campaign, but could be used to deal with similar alignment problems in the future.

3.2.2 Leeds LP-LIF instrument

A similar behavior of the decay curves as observed for the Lille LP-LIF instrument was recognized for the Leeds LP-LIF instrument after the campaign. In the decay, a fast component was followed by a slower component rather than single exponential behavior. This behavior was also apparent during the zero decay measurements conducted with zero air for this campaign. As a consequence, the fit of the single exponential function was started after the fast section of the decay curve for the data evaluation (fit range between 150 and 400 ms) for low reactivity values ($k_{OH} < 10 \, \mathrm{s}^{-1}$). An accurate determination of the OH reactivity was more difficult for high reactivity values ($k_{OH} > 10 \, \mathrm{s}^{-1}$), when values became similar to the fast component of the decay. A single exponential function between 100 and 200 ms was fitted to the measured decay curve in this case.

Similar to the procedure that was applied to the data of the Lille LP-LIF instrument, zero air measurements were evaluated using the same fit ranges as for evaluating low and high reactivities values. A decay time of $(2.3 \pm 0.4) \, \mathrm{s}^{-1}$ was obtained for the low reactivity case. This is close to the real loss rate of OH in the instrument without OH reactants (instrument zero). A higher value of $(4.8 \pm 0.6) \, \mathrm{s}^{-1}$ was determined, if the fit range was shifted to an earlier start as for evaluating decays for high

reactivity values. These two values were subtracted as instrument zeros when either one of the fit ranges were used. The higher value acts as a correction for the overlap of the faster instrumental component and the OH decay due to chemical reactions. For decays taken on the 7 and 15 October 2015 when NO was present, a fit range between 105 and 150 ms was chosen giving an instrument zero of $(5.1 \pm 1.1)\,\mathrm{s}^{-1}$.

5 The difference between the revised data, when this evaluation scheme was applied, and the initially submitted data is mainly due to the higher instrument zero that was subtracted for $k_{\mathrm{OH}} > 10\,\mathrm{s}^{-1}$, so that these values are $2.5\,\mathrm{s}^{-1}$ lower than before. Deviations of the OH decay from single exponential behavior for conditions without OH recycling in the instrument were not observed in other field campaigns in the past. This correction is specific for the data from this campaign.

3.3 Direct OH loss rate measurement by flow tube technique with laser-induced fluorescence (PSU instrument)

10 The flow tube LIF instrument of PSU also measures the decay of OH radicals. In contrast to LP-LIF instruments, OH is continuously produced by water photolysis at 185 nm in this instrument using a pen-ray lamp with concurrent HO_2 production as in the CRM instruments.

In the PSU instrument, the reaction time is varied by a movable injector, which is used to change the distance between OH injection and the point of OH detection (Kovacs and Brune, 2001; Mao et al., 2009). The reaction time is calculated from 15 the velocity measured with a hot-wire anemometer and the known distance traveled for each position of the injector. Within each scan, more than 100 data points were used to calculate the decay. Finally, during normal operation in the field, the PSU instruments samples ambient air with a high flow rate ($>100\,\mathrm{liter/min}$). This exceeds the flow rate which can be consumed during operation of the SAPHIR chamber, therefore the PSU instrument had to apply a high dilution flow in this campaign. Only $20\,\mathrm{liter/min}$ were sampled from the chamber, to which $80\,\mathrm{liter/min}$ of high purity synthetic air provided by the SAPHIR 20 air supply system was added. The dilution factor was determined from monitored flow rates and was verified in several tests during the campaign, in which the ratio of flows was varied. Using a dilution flow has two drawbacks. First, the calculated OH reactivity is very sensitive to the exact ambient and dilution flows. Second, any error in the instrument zero decay due to wall loss or trace impurities in the dilution air, is amplified by the ratio of the dilution flow to the ambient flow, in this case a factor of 5. Thus the typical limit of detection of $0.5\,\mathrm{s}^{-1}$ becomes $2.5\,\mathrm{s}^{-1}$.

25 As for the CRM instruments, measurements by the PSU instrument can be affected by OH recycling from the reaction of ambient NO with HO_2 that is concurrently produced with OH by water photolysis. The correction of OH recycling in the PSU instrument is based on correcting each point in the decays for the recycling calculated from measured NO and HO_2 before applying the fit to determine the OH reactivity (Shirley et al., 2006).

Changes made after the first data submission in the data by the PSU instrument were mostly smaller than $\pm 10\,\%$. These small 30 changes were due to improvements in the data evaluation algorithms that were made between the first and final submissions. These included improvements in the procedure, how data measured from instruments that were used for the corrections were synchronized to the OH reactivity measurements, and refinement of instrument parameters such as air velocity and location of the injector.

In addition, the change in the correction procedure for OH regeneration due to the reaction of HO_2 with NO led to the final data being 2.5 times higher than the first data submission at the highest NO mixing ratios on 07 October 2015. Initially a new optimization fitting procedure was developed and used for the first data submission, but laboratory and modeling studies showed that the method in Shirley et al. (2006) was superior and less uncertain. Thus, the method in Shirley et al. (2006) was used for the revised final data submission.

These changes were specific for this campaign because the instrument was not exactly the same instrument as used in previous and future campaigns. It was assembled from parts of the original PSU instrument and parts (mainly the laser system for the OH detection) provided by Max-Planck-Institute for Chemistry in Mainz and the University of California, Berkeley.

3.4 Direct OH loss rate measurement by flow tube technique with chemical ionization mass spectrometry (DWD instrument)

The measurement scheme of the CIMS instrument by DWD is similar to that of the flow tube LIF instrument by PSU. However, only one reaction time is currently realized to measure the OH decay (Berresheim et al., 2000; Muller et al., 2017). Excess OH ($10^8 \, cm^{-3}$) is produced by water photolysis in front of the flow tube with concurrent production of HO_2. The reduction of its concentration by reacting with ambient OH reactants is measured at two set time periods. This is achieved by terminating this reaction by chemical conversion of OH after a specific reaction time. For this purpose, a high concentration of sulfur dioxide is added at two injection points, so that OH is converted to sulfuric acid. After the OH titration, a high concentration of propane is injected to scavenge any OH present. The injection of sulfur dioxide is alternately switched between these two points. The reaction time is determined by adding known amounts of OH reactivity (e.g. propane) in front of the flow tube. OH wall losses on the flow tube are quantified by using humidified synthetic air. If the OH lifetime in the instrument is on the order of the travel time between the two injection points, no reasonable measurement is possible. In the current set up, an upper limit of OH reactivity values of $40 \, s^{-1}$ is achieved.

Additionally, measurements by the CIMS instrument can also be affected by OH recycling from the reaction of ambient NO with HO_2 that is concurrently produced with OH by water photolysis. Corrections for OH recycling in the CIMS are based on laboratory characterization at the Hohenpeissenberg station (ambient pressure $\sim 900 \, hPa$). An empirical function corrects for the systematic underestimation seen in CIMS OH reactivity measurements, which is both dependent on the magnitude of OH reactivity and the levels of NO present. The function has been derived for propane, isoprene and ethene for NO concentrations up to $15 \, ppbv$ (Muller et al., 2017). Under the assumption that any complex mixture in the SAPHIR chamber behaves like the three OH reactant mixture above, the NO correction was applied to the SAPHIR campaign dataset for k_{OH} larger than $2.5 \, s^{-1}$. The fit function optimized for OH reactivity up to $40 \, s^{-1}$ and NO ranging from 0 to $15 \, ppbv$ leads to an systematic overestimation of OH reactivities below $2.5 \, s^{-1}$ (Fig. 3 in supplementary information), not representing laboratory observations. Therefore no correction is applied for $k_{OH} < 2.5 \, s^{-1}$. The OH recycling efficiency is partly dependent on the reaction time between the two injections zones. As the NO correction was determined in the laboratory at Hohenpeissenberg Observatory at a pressure of $900 \, hPa$, an uncertainty of $+10 \, \%$ exists for its application at the SAPHIR chamber, as a result of lower flow rates (i.e. longer reaction time in CIMS) at $1000 \, hPa$.

In addition to ambient NO, the CIMS measurements were influenced by an NO impurity in the SO_2 cylinder, leading to the presence of $0.14\,ppbv$ NO in the CIMS flow tube at all times in this campaign. The presence of the NO impurity became evident from the inspection of the CO experiments (07 and 15 April 2016) where a systematic, repeatable underestimation in OH reactivity was found for reactivities above $20\,s^{-1}$. Therefore, an NO correction function was applied to the whole dataset also for experiments without NO in the chamber, e.g. in experiments with monoterpenes (13 April 2016) and sesquiterpenes (15 April 2016).

The DWD CIMS instrument is a relatively new instrument that had only been used in a remote environment at the monitoring station at Hohenpeißenberg. Therefore, the correction procedure had been developed for chemical conditions experienced in this campaign and were further refined after the first data submission. They would also be required, if the instrument measured in similar environments.

The wall loss of OH in the instrument and the time that the air traveled between the two titration points were initially determined from zero air phases of the experiments on each day. In order to provide data which are fully independent from the experiments, measurements were revised after data of the other instruments were known. The parameters were determined by an external flow tube with propane and synthetic air concentrations only once before the start of the campaign. This resulted not only in a constant change in the data due to the change of the zero decay time (wall loss) but also a scaling of data due to the change of the calculated reaction time (Table 1 in the supplementary information). Final data are on average $10\,\%$ lower than initially submitted.

4   Results and Discussion

A summary of OH reactivity measurements of all instruments together with calculated OH reactivity is shown in Fig. 1 to 4 and results are discussed in detail in the following sub-sections. For comparing data, the calculated reactivity is taken as the reference value, if no oxidation products were formed during the experiment. In all other cases, one of the LP-LIF instruments (FZJS) is taken as reference. This instrument was chosen because its measurements have a high precision and time resolution. Regression lines were determined using the fitexy procedure by Press et al. (1992). This method takes into account the measurement errors of both instruments and is symmetric, i.e. the fitted parameters are independent of which of the two instruments is assigned to be the dependent or independent variable.

4.1   OH reactivity measurements with zero air

Ultra-pure air was present in the chamber in the beginning of each experiment. As discussed above, it can be assumed that there was no OH reactivity present in this case. For normal operation of the LIF instruments, ozone and water vapor need to be present. A small contamination from OH reactants could appear during the humidification process of the chamber air. Measurements from previous experiments in the chamber indicate that OH reactivity introduced together with water is most often below the limit of detection of the reactivity instrument (approximately $0.2\,s^{-1}$, e.g. Fuchs et al. (2013)), but always less

than $1\,\mathrm{s}^{-1}$. This is likely due to either contaminants in the water or contaminants coming off the Teflon film of the chamber with increasing humidity. Therefore, these periods are ideal to test the instrument zeros and the precision of measurements.

If an instrument zero needed to be taken into account, it was independently determined from the zero air phase of the experiments for all instruments except for the MDOUAI CRM instrument. The instrument zeros were typically measured on a daily basis. No systematic change of the value was observed over the course of the campaign for these instruments. The instrument zero of the Leeds LP-LIF instrument was determined only once at the end of the campaign and the zero of the CIMS instrument once at the beginning of the campaign. The same air supply as for the chamber was used in these experiments except for the CIMS, for which bottled air was used (Linde, purity $99.999\,\%$). The derived values were used to correct all data. No instrument zero is expected for the CRM instruments (except for the contamination in the MDOUAI instrument), because only differences between measurement modes are used to calculate the OH reactivity.

Figure 5 shows the histogram of measurements during all zero air parts in the two campaigns. A Gaussian fit function is fitted to the distribution in order to determine a potential bias in measurements and to estimate the precision of measurements (Table 5). Overall, the distributions of zero measurements give a Gaussian shape. If all data are put together, none of the instruments exhibit a significant bias. Some exceptions are observed for specific experiments for some instruments. This result also demonstrates that no significant OH reactivity was present during these parts of the experiments.

Partly due to the small number of data points, the distribution is noisier for the measurements by the CRM instruments compared to distribution for the LP-LIF instruments. No bias of the MDOUAI instrument can be determined because of the use of zero air phases of experiments to determine an instrument zero (see above). The bias in the other two CRM instruments varies between experiments (Fig. 1); the day-to-day variability is between ± 3 and $\pm 5\,\mathrm{s}^{-1}$ with maximum values of $\pm 10\,\mathrm{s}^{-1}$.

A small bias is also observed in a few experiments for measurements of the Leeds LP-LIF instrument (smaller maximum value around $2\,\mathrm{s}^{-1}$). A small positive bias of approximately $2\,\mathrm{s}^{-1}$ is also seen in the PSU measurements after 13 October 2015 for unknown reasons. However, this change is likely affected by the amplification of zero variability and errors due to the dilution procedure that was used.

The width of the distribution can only be regarded as an upper limit for the precision because of the deviation of zero measurements from a Gaussian distribution for some instruments (Table 5). Alternatively, the width of the distribution was calculated for a distribution of data after subtraction of the bias observed for each instrument for each individual experiment. The width of this distribution gives a more realistic estimate of the precision of measurements (Fig. 5 in the supplementary information). The widths of the corrected distributions for the CRM instruments give a precision of approximately $2\,\mathrm{s}^{-1}$ at a time resolution of 10 (LSCE-CRM, MDOUAI-CRM) or $15\,\mathrm{min}$ (MPI-CRM), slightly higher than the stated limits of detection of 1 to $1.5\,\mathrm{s}^{-1}$ (Table 1). The widths of the distributions give a precision between 0.1 and $0.3\,\mathrm{s}^{-1}$ for LP-LIF instruments at time resolutions between 30 and $160\,\mathrm{s}$ for the different instruments, and a precision of $0.4\,\mathrm{s}^{-1}$ for the CIMS (60 to $300\,\mathrm{s}$ time resolution) in agreement with their stated limits of detection (Table 1). The PSU flow-tube instrument gives a precision of $0.9\,\mathrm{s}^{-1}$. The precision in Table 1 of $0.5\,\mathrm{s}^{-1}$ is stated for normal operation of the instrument without the dilution and becomes $2.5\,\mathrm{s}^{-1}$ when corrected for the dilution amplification.

4.2 OH reactivity measurements in the presence of CO and CH_4

During several experiments, only CO and CH_4 were present in the chamber for the entire experiment or part of the experiment. These experiments were done in the dark, so that there was no photochemistry. The linearity of instruments and behavior for a chemically simple system can be investigated from these experiments.

As can be seen in Fig. 1 and 3, measurements of all instruments followed the expected changes (expressed by k_{calc}) due to the additions of CO and CH_4. In the tested range of up to $150\,s^{-1}$, the agreement is mostly very good. Measurements by the DWD instrument show a clear upper limit of measurable reactivity of $40\,s^{-1}$ as expected from the measurement principle (see above). Some instruments exhibited large transient deviations from the expected values (e.g., MPI on 06 October 2015 between 10:00 and 11:00), but agree otherwise well during the CO/CH_4 experiments.

Although linearity appears not to be a problem for all instruments even for exceptionally high reactivity values (Fig. 6 and Table 2 both in the supplementary information), the discussion of results focuses on OH reactivity values below $60\,s^{-1}$, which are more relevant for atmospheric measurements. Figure 6 shows the correlation of measured and calculated OH reactivities and Table 6 gives the result of a regression analysis for all periods, when only CO and/or CH_4 were present in the dark chamber. High linear correlation coefficients ($R^2 > 0.8$) are calculated for all instruments. A linear regression analysis gives slopes between 0.98 and 1.17 for most instruments. Errors of the fitted slopes were always smaller than 0.01, because the precision of data are higher than the scatter of data around the regression line. These results demonstrate the ability of instruments to measure the correct reactivity values. Only the regression analysis for one CRM instrument (LSCE) gives a higher slope of 1.31 mostly due to measurements during the first experiment, whereas better agreement (Table 6) is achieved in other experiments. The larger deviation for this instrument is likely due to the correction for deviations from the pseudo-first order behavior (Table 4). This was determined from characterization measurements with a mixture of isoprene and propane, which might not well represent chemical conditions with only CO.

Although the slopes of the regression lines indicate on average good agreement of measurements for these chemical conditions, the scatter in the correlation plots (Fig. 6) is considerably different for the instruments. The time series in Fig. 1 and 3 show that the scatter in the correlation plot is caused by statistical noise, and for some instruments by irregular systematic deviations pointing to instrumental instabilities. The mean of the relative absolute difference between measured and calculated OH reactivity is 32 to 48 % for CRM instruments, 19 % for the PSU LIF instrument and between 8 to 11 % for LP-LIF and the CIMS instruments. If the PSU instrument was operated like in the field without the large dilution flow, measurements would have scattered significantly less (at least a factor of 5). Thus, for the instruments as configured for this comparison study, the PSU LIF and LP-LIF instruments appear to have the highest measurement precision.

4.3 OH reactivity measurements in the presence of isoprene, MVK, MACR and OH reactants in urban environments

In a second set of experiments, chemical conditions included volatile organic compounds, NO_2, and CO (Table 3). The most abundant biogenic species, isoprene, and OH reactants that are representatives for alkenes and aromatic compounds found in

urban environments (1-pentene, o-xylene, toluene) were tested. Oxygenated VOCs from isoprene oxidation (MVK and MACR) and acetaldehyde were present in separate experiments in 2015 (Fig. 8). In 2016, these species were present in experiments together with isoprene and the urban OH reactant mixture.

Similar results are obtained for isoprene and urban OH reactants. Because these experiments partly included oxidation products that were not measured by instrumentation at the chamber, measurements by the LP-LIF FZJS instrument are taken as the reference value. Measurements of this instrument differ less than 10 % from calculations using measured OH reactant concentrations. This difference is smaller than the 1σ accuracy of the calculation, so that results would not significantly differ, if calculated OH reactivity was used.

For most instruments (except for LSCE CRM and DWD CIMS instruments), the agreement between measurements found for these chemical conditions is about as good as for the experiments with only CO and CH_4 (Fig. 7). High linear correlation coefficients between measured and calculated reactivity values are obtained ($R^2 > 0.80$) and slopes of the regression lines are 0.94 and 1.07 showing good absolute agreement (Table 6).

The performance of the LSCE instrument is better than in the experiments with only CO/CH_4, but measurements are lower than the reference in this case, whereas measurements are higher in the CO/CH_4 case. As discussed above, the correction for deviation from pseudo-first order kinetics (which is based on a characterization with propane and isoprene standards) might better represent chemical conditions during the experiment with alkenes, aromatics and isoprene compared to the CO/CH_4 case. In general, this issue can cause a variability in the agreement between measured and calculated reactivity in this campaign. This indicates that a more intensive characterization of this correction is required for the specific chemical conditions, specifically, if individual OH reactants are studied.

Measurements by the DWD CIMS instrument give larger deviations from calculated reactivity in these experiments compared to results found in the CO/CH_4 case. The experiments on 09 and 11 April 2016 started with high reactivities of about $40\,s^{-1}$, but only 60 % is measured by the DWD CIMS instrument (Fig. 3). The agreement improves when k_{OH} decreases. For values below $10\,s^{-1}$, the measurements well agreed with calculated reactivities. The exact behavior of the relationship between measured and calculated reactivity changed for periods of the experiments with different chemical conditions. In addition, an increase in the measured reactivity with increasing water vapor concentration after starting humidification of the chamber air is observed in these experiments. This is less obvious in other experiments (see below).

Part of this large discrepancy could be the result of an instrumental instability, which was seen as an intermittent increase in noise in the CIMS reactant ion counts (NO_3^-) from 09 April 2016 onwards and coincided with the periods deviating from the FZJS instrument observations (see supplementary information). This could be relevant because the reactant ion count is used to normalize the HSO_4^- counts, thus obtaining the equivalent OH concentration. At high OH reactivities, when OH signals are smaller, higher noise in the (comparatively) large reactant ion concentrations could thus affect the resultant k_{OH} estimation. The exact reason why there was an increase in noise in reactant ion concentrations remains unclear. Additionally, these two experiments have in common the illumination of the chamber by sunlight and presence of NO_2 in the second part of the experiments. Interestingly, measured and calculated reactivity agree better in these parts of the experiment compared to the first parts. However, there is no obvious reason why these conditions would impact the measurements of this instrument.

Chemistry occurring in the inlet system may impact the OH concentration for the more complex chemical composition of air. In the presence of NO (see below), any unaccounted OH recycling would affect the accuracy of measurements.

In the 2016 experiments, the relationship between measured and calculated reactivities does not change when oxygenated VOC (MVK, MACR and acetaldehyde) were present compared to the part of the experiments, when only the parent VOC was present. This can be seen in the time series of experiments on 09 and 11 April 2016 (Fig. 3). In 2015's campaign, the impact of the presence of these compounds on the instruments were tested in a separate experiment (16 October 2015, Fig. 2). Because the compounds were consecutively injected, only the observed change in the measured OH reactivity for each instrument is calculated for the analysis (median of 20 min of measurements before and after the injection). This value can be compared to the expected change of the reactivity that is calculated from measured reactant concentrations (Fig. 8).

Measurements of LP-LIF instruments are not affected by these species and agree with the change of calculated reactivity. Also the flow tube LIF instrument by PSU and the CRM by MPI give similar values within 10 to 20 %. The change in reactivity measured by the LSCE instrument agrees in the case of MVK and MACR, but is less in the case of acetaldehyde. Changes measured by the MDOUAI CRM instrument are up to a factor of three lower than observed by the other instruments. Losses on surfaces in the inlet system may explain part or all of the discrepancy. Both instruments used an additional pump with Teflon surfaces in their inlet system (see also the discussion for monoterpenes / sesquiterpenes).

The largest differences are seen in the presence of acetaldehyde for the MDOUAI CRM instrument. In addition, measurements by the LSCE and MDOUAI instruments are more variable compared to those by the other instruments as indicated by the large difference between 25 to 75 percentile values in this case. The presence of oxygenated VOCs may cause additional complications in the reaction system in the CRM that impacts the OH concentration. The oxidation of aldehyde species by OH proceeds by H-atom abstraction from the aldehydic group, leading to the formation of acyl peroxy radicals, $RC(O)O_2$. For instance, the oxidation of acetaldehyde will lead to the formation of the acetyl peroxy radical, $CH_3C(O)O_2$, with a yield of approximately 95 % (Cameron et al., 2002; Butkovskaya et al., 2004). The reaction of acyl peroxy radicals with HO_2 is known to efficiently recycle OH in the atmosphere. For the acetyl peroxy radical, Dillon and Crowley (2008); Winiberg et al. (2016) recently reported an OH yield of 0.5. Also one reaction pathway in the reaction of MVK with OH forms an acyl peroxy radical that leads to OH reformation in the reaction with HO_2 with a high yield of 0.64 (Praske et al., 2015). These recycling mechanisms can act as a secondary source of OH in CRM instruments, which in turn can mask a fraction of the OH-reactivity from aldehyde species for these instruments, leading to a negative bias. Results of model calculations and laboratory investigations (Fig. 4 in the supplementary information) performed for the MDOUAI instrument confirm that the OH reactivity of acetaldehyde is underestimated by this instrument, which is consistent with observations during the 16 October 2015 experiment (Fig. 2), when acetaldehyde was first introduced in SAPHIR. However, Fig. 2 shows that the two other CRM instrument (LSCE and MPI) are less (or not) impacted by OH recycling from $CH_3C(O)O_2+HO_2$. These different behaviors are not well understood and need more investigations.

It is noteworthy that concentrations of acetaldehyde and other aldehydes in the atmosphere are typically smaller than in the experiment in this campaign, but can be a significant fraction (10 to 20 %) of the total reactivity (e.g. Fuchs et al., 2017). The

maximum error that is caused by the underestimation of the total reactivity measurement by the CRM instrument would be less than 13 % if results of the experiment of this campaign are extrapolated to atmospheric conditions.

4.4 OH reactivity measurements in the presence of monoterpenes and sesquiterpenes

The third type of chemical conditions tested in the campaigns was the presence of terpenes. This was done either by injecting a mixture of monoterpenes, a sesquiterpene or by flushing real plant emissions into the SAPHIR chamber. These experiments also included ozonolysis reactions of terpenes. Maximum reactivities ($25\,\mathrm{s}^{-1}$) were lower than in other experiments in 2015. Oxidation products of the ozonolysis reactions were not measured, so that it is expected that calculated reactivities are underestimating the real reactivity. Therefore, one of the instruments (LP-LIF FZJS) is taken as reference for the comparison of the measurements. As seen in the correlation plots (Fig. 9) and the results of the regression analysis for data without the presence of ozone (Table 6), differences between measurements of the LP-LIF instruments and the other instruments are largest in these experiments compared to the other experiments.

High linear correlation coefficients are obtained ($R^2 > 0.96$) and slopes of the regression lines between 0.96 and 1.08 are calculated for the LP-LIF instruments. No systematic change in the relationship of measurements is observed whether ozone and hence ozonolysis products are present or not. Similarly, measurements between LP-LIF instruments agree in the presence of the sesquiterpene (with and without the presence of ozone and ozonolysis reaction products). Because this experiment started with the addition of other OH reactants (Fig. 2), only the measured difference due to the injection of $\beta-$caryophyllene that is observed by each instrument is compared (Fig. 10) as done for the oxygenated VOCs (see above).

Measurements of the flow-tube LIF instrument (PSU) varied more with respect to the reference measurements in the presence of monoterpenes and sesquiterpenes compared to the other chemical conditions. The level of agreement varies among the three experiments (9, 14, 16 October 2015): when the monoterpene mixture was injected, measurements by the PSU instrument are 10 to 15 % lower than measurements by the FZJS instrument, but they are 20 % higher, when plant emissions were transferred into the chamber. During the continuous transfer small inhomogeneities cannot be fully excluded, but the discrepancies between measurements remains also after the injection and oxidation part of the experiment. The relationship does not depend on the presence of ozone in these two experiments. In the third experiment, changes in the OH reactivity measured by the PSU instrument due to the increase of sesquiterpene concentration (and ozonolysis products) are up to 40 % smaller than the changes observed by the LP-LIF instruments (Fig. 10). The higher and lower values observed in these experiments may not be related to the chemical conditions, but instead instrument problems. This is indicated by higher values of the PSU instrument compared to measurements by the other instruments in nearly all experiments after 13 October independent on the chemical conditions (Fig. 7, 9). Difficulties in maintaining consistent operation of the laser and the electronics driving the movable OH source could have led to much of this variability. As a result, this comparison exercise probably does not represent the capability of the PSU instrument to measure OH reactivity in forest environments.

Measurements by FZJ LP-LIF and DWD CIMS instruments in the experiments with terpenes (Fig. 9, 11) well agree during the first experiment with monoterpenes (13 April 2016, Fig. 4) and only a small underestimation is seen during the second experiment (14 April 2016). Though no obvious explanation can be provided why e.g. the CIMS underestimated OH reac-

tivity compared to LP-LIF up to $12\,\mathrm{s}^{-1}$ on 14 April 2016, the CIMS instrument performance might have been influenced by unidentified internal chemical reactions. This also corresponds to observations in the presence of isoprene, MVK, MACR or a mixture of urban OH reactants (see above).

Because NO was present as impurity in the CIMS sulfur dioxide titration gas mixture (see above and Table 4), a NO correction function was applied also in experiments with monoterpenes (13 April 2016) and sesquiterpenes (15 April 2016). This could explain some of the smaller, but systematic differences compared to LP-LIF measurements (Muller et al., 2017).

Similar to the CIMS instrument, the agreement between CRM and LP-LIF instruments is worse in the presence of monoterpenes and the sesquiterpene compared to other experiments. Lower linear correlation coefficients (R^2) between 0.48 and 0.72 and a higher scatter of data with relative mean absolute residuum values between 0.34 and 0.45 (numbers only for periods without ozone) are observed. The agreement is even worse during the ozonolysis parts of the experiments, when CRM instruments measure values that are up to five times smaller than measurements of the LP-LIF instruments. Similar results are seen in the experiment with the sesquiterpene (Fig. 10). In all these cases, the level of agreement varies among the CRM instruments and the specific experiment, but measurements tend to be significantly smaller than those of the other instruments.

The residence times in the sampling lines of the CRM instruments were generally longer (5 to 6 s) compared to the sampling lines of the other instruments (0.5 to 4 s) and the volume to surface ratio was lower, because CRM instruments used 1/4" OD PFA tubing in 2015. In addition, two CRM instruments (MDOUAI and LSCE) used a sampling pump with Teflon surfaces to introduce the sample into the CRM reactor. Oxygenated and low volatility (monoterpene and sesquiterpene) species may adsorb on these surfaces and the pump may have therefore played a role in the underestimation seen for these instruments. One instrument (MPI-CRM) used a heated inlet line. Other instruments used up to 1" OD PFA or Silconert-coated stainless steel tubing (Table 1). Results (Fig. 9, 10) show that MPI measurements are partly significantly higher than those of the LSCE and MDOUAI instruments, supporting that the underestimation of the LSCE and MDOUAI CRM instruments could be partly due to a loss of OH reactants in the sampling system (unheated inlet line+pump). However, an impact of the monoterpene or sesquiterpene chemistry on the CRM measurements cannot be ruled out.

In the experiments with terpenes in 2016 (Fig. 9, 11), a better agreement between measurements by the MPI CRM and FZJS LP-LIF specifically for the experiment with $\beta-$caryophyllene is found compared to the experiments in 2015. The reason for this improvement is not clear, but could be related to the larger diameter inlet tube used in 2016 compared to 2015 (Table 1) supporting the potential influence of losses in the inlet system for these compounds.

4.5 OH reactivity measurements in the presence of NO

The presence of NO can affect measurements of the OH reactivity in all instruments due to the recycling of OH by the reaction of HO_2 with NO that is contained in ambient air (see above). These effects are amplified, if OH is produced by water photolysis, because HO_2 is concurrently formed with OH.

In 2015, the NO concentration was increased stepwise (up to $120\,\mathrm{ppbv}$) in the presence of CO on 07 and 15 October 2015 (Fig. 1). NO was also present in the two experiments with urban OH reactants (12 and 13 October 2015). Figure 12 shows the dependence of the relative difference between measured and calculated reactivity on the NO mixing ratio in these experiments.

In 2016, the presence of NO was tested in two experiments in combination with the presence of pentane and in the urban OH reactant mixture (08 and 12 April 2016, Fig. 13). Due to the lack of OH reactant concentration measurements on 08 April 2016, measurements performed by the FZJS instrument are taken as the reference to analyze the impact of NO on the performance of the other two instruments (Fig. 13).

5   Discrepancies between calculated and measured reactivity are mostly within the range of differences observed in other experiments for LP-LIF instruments. For NO mixing ratios higher than $20\,ppbv$, median values deviate up to $20\,\%$ from calculated reactivities. The OH production rate from recycling reactions is within the range of the OH destruction rate in the case of CO at highest NO mixing ratio in the experiment on 07 October 2015. Nearly, all LP-LIF instruments applied a bi-exponential fit function in this case except the Leeds LP-LIF instrument. Only FZJM/FZJS instruments applied a bi-exponential

10 fit to measurements on the second experiment with high NO (15 October 2015). However, separation of OH reactivity from OH recycling by applying a bi-exponential fit function might still lead to some systematic errors for the experiment with high NO in the presence of only CO because of the high OH recycling rate that was higher than for typical atmospheric conditions. Therefore, the faster decay rate could deviate from the OH reactivity. The scatter in the measurements increases with increasing NO mixing ratio as expected from the lower precision of measurements for high reactivity values.

15   As with all other LP-LIF instruments, OH reactivity from the Leeds LP-LIF well agrees with calculated OH reactivities below $20\,ppbv$ NO. However, values are increasingly lower for higher NO mixing ratios. The lower values of OH reactivity for NO higher than $20\,ppbv$ are caused by the application of a single exponential fit to the OH decay data rather than using a bi-exponential fit as used by other LP-LIF groups. A similar behavior is achieved, if a mono-exponential fit is applied to measurements by the Lille and FZJS/FZJM LP-LIF instrument during the experiment with CO and NO (07 October 2015). A

20 bi-exponential fit, although giving an OH reactivity closer to the calculated value for this particular experiment, is not necessarily the correct function to apply for fitting of atmospheric data, and so was not used by Leeds (even though a bi-exponential fit returns a larger value of OH reactivity). Model simulations under relevant conditions indicate that a bi-exponential fit can return an OH reactivity that is greater than the true value. To fit the data more rigorously requires a modeling approach, similar to that applied when OH recycling was observed in a laboratory kinetics study (Onel et al., 2014). Hence, although application

25 of a bi-exponential fit improves the agreement with the calculated value for this experiment, caution is needed in applying it to atmospheric data at high NO where conditions could be different.

    Measurements by the PSU LIF instrument, which also uses water photolysis as an OH source, show a tendency to underestimate reactivity values with increasing NO mixing ratios. The maximum median of the relative difference is $20\,\%$. This difference was significantly reduced from $70\,\%$ to $20\,\%$ in the final data compared to the data submitted before reactivity mea-

30 surements from all groups and OH reactant concentrations were made available. The correction procedure for the presence of NO was changed later from a new procedure to the one described in Shirley et al. (2006) (see above).

    Measurements by the MDOUAI and LSCE CRM instruments do not exhibit a clear trend in the relative difference between measured and calculated reactivity with NO. In contrast, measurements by the MPI CRM instrument give lower reactivity values compared to calculated reactivities with increasing NO mixing ratios in both campaigns in 2015 and 2016 (Fig 12, 13).

35 Measurements by all CRM instruments were corrected by applying an empirical correction function. The magnitude of the

correction is on the order of the OH reactivity values (Table 4) making results very sensitive to any systematic error in the correction procedure. The differences in the corrections needed for each CRM instrument emphasize the necessity for a careful characterization of the instrument.

In the campaign in 2016, the relative difference between measured reactivity by the DWD CIMS instrument and the reference (FZJS LP-LIF instrument) is small for NO mixing ratios lower than 5 ppbv, but increases with increasing NO mixing ratio (Fig. 13) to up to a factor of 1.3 (median value) for 10 to 20 ppbv NO.

This difference demonstrates that the correction applied for the CIMS measurements leads to systematic errors for NO mixing rations larger than 5 ppbv in particular for the urban mixture (12 April 2016). The chemical composition does seem to play a role (Table 3); the correction in the pentane experiment applies well in contrast to partly producing inaccurate results in the urban mix experiment. The strength of OH recycling by the reaction of HO_2 with NO is also dependent on the CIMS internal abundance of HO_2 which was not measured during the SAPHIR campaign. Also, the correction term can become rather large for high OH reactivity and large NO concentration (up to $30\,s^{-1}$) which illustrates the limit of the instrument in its current configuration (Fig. 4).

4.6 Influence of humidity on OH reactivity measurements

In experiments in 2015, humidity was similar in most experiments and humidity was only systematically varied in one experiment on 06 October 2015 (Fig. 1). In contrast, water vapor concentrations were highly variable in experiments in 2016 because of the high dilution flow that was required. Figure 14 shows the dependence of the relative difference of measurements by the instruments (taking measurements by the FZJS instrument as reference) for all experiments, in which an overall good agreement is observed (CO, pentane). A clear trend to over-predict OH reactivity with increasing water vapor can be seen for measurements by the MPI CRM. This trend is consistent with lower measurements at the lowest water mixing ratios observed in the experiment on 06 October 2015 (Fig. 1, Fig. 7 in the supplementary information). In contrast, the results from 06 October 2015 do not indicate that the other CRM instruments are affected in the same way by water vapor. No clear trend with water vapor is observed for the LSCE and MDOUAI instruments.

Some changes in the relationship of the CIMS instrument with the LP-LIF instrument are seen after water vapor additions in some experiments in 2016 (for example 09 and 11 April 2016 Fig. 3). On 09 and 11 April 2016, the CIMS instrument showed an instrumental instability before the water addition, which could explain some changes in the relationship (Fig. 2 in supplementary information). No systematic trend in the entire dataset is observed (Fig. 14). Also on 15 April 2016 CIMS measurements deviate from observation of the LP-LIF instrument, when the humidity was increased after the addition of sesquiterpenes, but deviations could be due to errors in either one of the instruments. The LP-LIF instrument observed an increase of OH reactivity, whereas the decreasing trend of the CIMS measurements does not change. The increase observed by the FZJS LP-LIF instrument could be due to desorption of sesquiterpenes inside the instrument, but could also be due to desorption from the chamber wall increasing the reactivity in the chamber. However, the decrease observed by the CIMS instrument would be consistent with the dilution of trace gases in the chamber. Both instruments better agree again after the injection of ozone, when sesquiterpenes have reacted away.

In the CRM and CIMS instruments, the concentrations of OH and HO_2 depend on the water vapor concentration as they are produced together by water vapor photolysis. In CRM instruments, corrections are applied, when the water vapor concentration changes between the different measurement modes which are required to calculate the OH reactivity (Table 4). Also a humidity dependence in the detection sensitivity of pyrrole is taken into account. Fast changes in the water vapor concentrations, for example like during the humidification procedure of the chamber air, can therefore cause systematic errors. However, systematic differences are observed on a longer time scale than the duration of the humidification ($<30\,\mathrm{min}$) in these experiments. Humidity dependent memory effects in the inlet system could be the reason for this behavior. Another possibility could be that observations are related to changes in the OH and HO_2 concentrations that depend on the water vapor concentration. OH recycling processes and correction factors depend on the radical concentrations and chemical conditions. For experiments in 2015, the dependence of the correction factor for the deviation from pseudo-first order kinetics was not well characterized for the low OH concentrations at low water vapor concentrations, so that the deviation from calculated OH reactivity might be due to a systematic error in this correction for these conditions.

Further investigations will be necessary to understand the exact influence of the water vapor concentration on the OH reactivity measurements by the MPI CRM instrument and the reason for the instrumental instability of the CIMS instrument. It cannot be fully excluded that the observed effects are related to the experimental procedure of humidifying the chamber air that leads to a relatively fast change of the water vapor concentration.

5 Previous comparisons

Two comparisons of OH reactivity instruments were performed in the past. In one campaign, the MDOUAI CRM and the Lille LP-LIF instruments measured together on the campus of the University of Lille in October 2012 (Hansen et al., 2015). Either a complex mixture of VOCs or oxygenated VOCs or ambient air was sampled. Experiments with synthetic mixtures of hydrocarbons and OVOCs indicated that the CRM instrument was underestimating the reactivity by 39 and 53 %, respectively, while the LP-LIF measurements were in agreement with the calculated reactivity values within their uncertainty. The discrepancy was attributed to the photolysis of aromatic compounds and oxygenated VOCs in the CRM instrument. This effect is expected to be insignificant in this campaign, because the lamp intensity in the MDOUAI instrument was lowered as a result of the campaign in 2012.

In the present study, a good agreement with calculated reactivity is found in experiments with urban OH reactants including aromatic compounds. This confirms that photolysis processes observed in 2012 for the MDOUAI CRM no longer play a role.

The comparison between the two instruments during ambient measurements showed that the CRM measurements were lower than the LP-LIF measurements by 22 % on average. No dependence in the agreement of the MDOUAI CRM and Lille LP-LIF on the presence of NO was observed in 2012, consistent with results in this campaign for these instruments (Fig. 12). Similar to the results of this campaign, the accuracy of the determination of the instrument zero of the LP-LIF instrument limited the accuracy of measurements.

Measurements of the MDOUAI CRM were also compared to measurements by the LSCE CRM during a field campaign at a remote site in France in summer 2013 (Zannoni et al., 2015). Both instruments sampled either ambient air or emissions from enclosed plants. Measurements by both instruments agreed well overall in that campaign (the regression yielded a slope of 0.96), but the linear correlation coefficient R^2 was only 0.54 for reactivity values below $50\,\mathrm{s}^{-1}$ because of the large scatter in the data at OH reactivity below $50\,\mathrm{s}^{-1}$. This is consistent with results in this campaign, where measurements from these two instruments were often similar, but also differed by 20 to 50 % in some experiments (Fig. 1 and 2).

The MDOUAI instrument was operated under similar conditions than during the present study, i.e. with a long sampling line and a pump in the inlet system. The inlet system of the LSCE instrument did not include a pump like in the present study and consisted of sampling Teflon line with small diameter (1/8"-OD) and a PTFE filter. It is likely that losses of low volatility compounds during sampling impacted the previous comparison, similar to observations in the present study.

6 Summary and conclusions

Measurements of OH reactivity were compared in experiments performed in the atmosphere simulation chamber SAPHIR in two campaigns in 2015 and 2016. All instrument types presently used for atmospheric measurements took part in one or both of the campaigns. A few additional instruments exist worldwide (e.g. Yang et al., 2016), but they are similar to instruments that participated in these campaigns.

6.1 Summary of findings

Not only were many measurements successfully performed in these campaigns, but also a number of findings led to an improvement of the data quality already during the process from measurements to the final data:

- An ozone dependent background signal was found for measurements of the MPI CRM.

- Application of the correction of measurements due to the deviation from pseudo first-order conditions in CRM instruments by empirical correction factors (Michoud et al., 2014) is recommended.

- Misalignment of the photolysis laser beam in the LP-LIF instruments can lead to a complication in the data evaluation procedure.

These results will also improve the precision and accuracy of measurements in the future.

The findings of the comparison of the final dataset are:

- Measurement techniques are capable of measuring OH reactivity for a range of chemical conditions that are relevant for ambient air measurements, but with different levels of precision and accuracy. Losses of OH reactants in inlet lines could be of importance.

- Measurements by LIF and CIMS instruments have a higher precision than CRM instruments leading to a limit of detection better than $1\,\mathrm{s}^{-1}$ at a time resolution of a few minutes. For chemically complex conditions, the scatter of the data is

within the range of 10 % for LP-LIF instruments and 10 to 20 % for the CIMS and the flow-tube LIF instruments. The precision of data from the flow-tube LIF instrument was reduced in this campaign compared to typical operation in the field due to the application of a high dilution flow. Measurements by CRM instruments exhibit a higher limit of detection of approximately $2\,s^{-1}$ at a time resolution of 10 to 15 min. The scatter of the CRM data for chemically complex conditions ranges from 17 to 45 % (mean relative difference between measurements and linear regression with reference values). Additional work is needed on the CRM technique to improve the measurement precision at a level closer to that observed for other instruments.

– Biases in the measurements by the LP-LIF instruments are lower than their limit of detection with a few exceptions in some experiments for the Lille and Leeds instruments. The instrument zero in the PSU instrument varied by $1.3\,s^{-1}$, but this value is amplified by the dilution factor of 5 that is not normally used in field measurements. The smaller number of data points for the CRM instruments makes conclusions about a day-to-day variability of a potential bias less accurate. However, the distribution of measurements during zero air measurements becomes more compact for the LSCE and MPI CRM instruments, if an offset is subtracted for each individual experiment (Table 5).

– Maximum absolute deviations of LP-LIF measurements from calculated reactivities or measured reactivities from the instrument taken as the reference (FZJS) are 12 % (mostly less than 5 %). Deviations are smaller than the accuracy of the calculation from measured OH reactant concentrations (Table 2). Results from this campaign demonstrate a high accuracy of LP-LIF instruments.

– The accuracy of CRM instruments varies with chemical conditions. Whereas measurements agree on average with calculated reactivities within 5 % (higher deviations for LSCE CRM 31 %), if only CO, CH_4 or pentane are present, deviation from measurements with the FZJS LP-LIF instrument are up to a factor of two for mixtures containing terpenic compounds. Also the scatter of data is larger in these cases. While the impact of OH recycling in the terpenic chemistry cannot be ruled out, losses of these compounds in inlet systems can explain the observed discrepancies. The transmission of low-volatility compounds such as terpenes and their oxidation products needs to be improved for CRM instruments Inlet systems used in this campaign partly differed from deployments in previous campaigns (for example the use of the additional pump in the inlet of the LSCE instrument), so that losses could have been different in campaigns in the past.

– Even in the presence of up to 120 ppbv NO, agreement with calculated reactivity within the accuracy of measurements and calculations is achieved for the MDOUAI and LSCE CRMs, whereas deviations of up to 50 % for the MPI CRM instrument and of a factor of 1.8 for the CIMS are observed. All these instruments applied large corrections to account for OH recycling from the reaction of HO_2 with NO. The variability in the accuracy of the correction emphasizes the need for a careful characterization of the instrument-specific operational conditions.

– Measurements by LP-LIF instruments are not affected as much as the other instruments by OH recycling reactions even for NO mixing ratios higher than 20 ppbv. In this case, a bi-exponential fit function to the OH decay curve rather than a single exponential fit improves the agreement with the calculated value of the OH reactivity. A bi-exponential function

was not applied to OH decays measured by the Leeds LP-LIF, so larger deviations were observed for NO mixing ratios higher than 20 ppbv. Although a bi-exponential fit to the data generated a closer agreement with the calculated value of the OH reactivity for the conditions of this particular experiment, it should be noted that it may not represent the best function to fit to atmospheric data at high NO (where the composition is different to this experiment), and careful thought needs to be given as to the optimum function to fit to the data.

- Measurements by the flow-tube LIF instrument give larger deviations ($\pm 20\%$) in the chemically more complex experiments compared to conditions with single, anthropogenic reactants (deviations $\pm 3\%$), although the flow-tube LIF measurements are likely affected by instrument issues related only to the instrument that was assembled for this comparison.

- Experiments in 2016 reveal a so far unrecognized effect of the water vapor concentration on measurements by the MPI CRM instrument (factor of two difference at 1% water vapor mixing ratio), although changes of the humidity and therefore radical concentrations are taken into account in the evaluation. The water vapor correction procedure might have not been applicable here, because humidity changes were faster than typically in the atmosphere. Water vapor was changed in only one experiment in 2015. Results do not indicate that the other CRM instruments are affected in the same way by water vapor.

- The accuracy of measurements by the CIMS instrument varied between experiments. Comparing to the calculated OH reactivity, an agreement is observed within the accuracy of measurements and calculations for the experiments with CO and pentane (deviation of the regression slope from 1:1 line of 13%). For the isoprene and urban reactant mixture cases, lower accuracy is observed with a deviation of the regression slope from 1:1 line of 27%. Contrasting to that, the regression slope is 1.01 for the monoterpene/sesquiterpene cases when measurements are referenced to the FZJS LP-LIF instrument. On some days a change in the relationship between measurements by the CIMS instrument and the LP-LIF instrument is observed with changing water. Overall, the variability in the level of agreement hints to instrumental instabilities.

6.2 Conclusions for future instrument operation and measurements in the past

Overall, the comparison demonstrates that OH reactivity measurements by LP-LIF instruments are precise and accurate for a wide range of atmospheric conditions. Instrumental parameters such as laser alignment and instrument zero are recommended to be regularly checked to achieve a high accuracy and to avoid additional complications in the data evaluation.

In this campaign, the flow-tube LIF instrument gives slightly less accurate and precise measurements compared to the LP-LIF instruments, which is related to the different operational conditions compared to previous campaigns. A different laser system was used and a high dilution flow was applied, which reduced the instrument performance. Had it been possible to use the field PSU instrument without dilution, it is likely that its precision and accuracy would have been similar to that of the LP-LIF instruments.

The OH reactivity scheme of the CIMS instrument is relatively new. This has only been deployed at the monitoring station Hohenpeissenberg so far, where OH reactivity values are typically small (2 to $10\,\mathrm{s^{-1}}$). Further improvements of the data quality for high NO conditions ($> 3\,\mathrm{pbbv}$) are needed to expand the device ability to measure in more polluted regimes.

The accuracy of the current observations depend on the quality of NO concentration measurements and the assumption that the OH decay obeys a single-exponential behavior. All OH recycling processes need to be well-characterized. Any deviation from these assumptions lead to systematic errors, and needs further investigations to capture also unknown complex mixtures of OH reactants under polluted conditions. Additional reaction times (injection points for SO_2 and propane) and concurrent measurements of RO_x and HO_2 concentrations could help characterize the OH recycling processes in unknown mixtures of OH reactants in the future.

While CRM instruments are less precise and accurate than other techniques, a reasonable agreement is usually observed between the CRM instruments and the other techniques for air mixtures containing simple compounds such as CH_4, CO, and isoprene, and for urban air mixtures containing anthropogenic hydrocarbons and NO_x. The correction factors, which depend on the exact instrumental conditions such as the OH, HO_2 and pyrrole concentrations in the reaction volume, are a potential source of systematic errors. In order to minimize these errors, the CRM operating conditions are such that the ratio of pyrrole to OH concentrations ranges from 1.7 to 2, so that corrections for operating under non-pseudo first order conditions can be within $10\,\%$ ($1\,\sigma$) for different air compositions (Michoud et al., 2015). The error associated to this correction needs to be propagated to the measurement uncertainty. The largest correction to apply is for the recycling of OH from HO_2+NO, which is only relevant for urban atmospheres. This correction can be on the same order of magnitude than the measured OH reactivity value and needs to be carefully characterized on each CRM instrument.

The level of agreement is degraded when low-volatility terpenoid compounds and/or their oxidation products are sampled. Although all CRM instruments use the same detection scheme and the same reaction tube, measurements differ between the three CRM instruments and also significantly differ from other LIF-based techniques. While OH recycling in the CRM reactor cannot be ruled out when these species are sampled, losses of OH reactants in inlet lines and sampling pumps (partly different than typically inlet lines in field campaigns) could have led to additional systematic errors in this campaign. The quality of the measurements depends on both the instrumental technique but also the procedure used to transfer the sample into the instrument. A high flow (short residence time) in the inlet lines and/or the use of inert inlet line materials like Silconert coated steel might help to reduce inlet line effects as indicated by the results of LP-LIF instruments. This improvement is a prerequisite to investigate whether the terpenoid chemistry inside CRM reactors can lead to an underestimation of ambient measurements.

The CRM method is a younger technique compared to the LP-LIF and flow-tube LIF method, but the number of instruments has quickly increased due to the commercial availability of detectors for pyrrole such as the PTR-MS instrument. Results of this campaign emphasize that careful instrument characterization for the specific operational conditions are required in order to achieve accurate measurements. Future work should focus on improving its performances in terms of precision and limit of detection. In addition, the accuracy of measurements would improve, if corrections could be lowered.

The results of this campaign demonstrate that all detection schemes that are currently applied for OH reactivity measurements give reasonable results for a range of chemical conditions which are relevant for ambient air measurements. These first

comprehensive comparison campaigns were conducted to assess the current performance of instruments. The results already led to the implementation of changes in some instruments to achieve better data quality. More work will be done in order to improve the instrument performance for issues that have been identified which currently limit the precision or accuracy of measurements. More comparison campaigns could help to further increase the trustability of measurements by conducting them in a formal, blind way and/or at even more realistic conditions with ambient air.

In the field, OH reactivity measurements are often used to identify unexplained reactivity from OH reactants that were not measured as individual species (Yang et al., 2016). Large unexplained reactivity (several $10\,s^{-1}$) was found in several field campaigns in biogenic environments such as the boreal forest in Finland (Nölscher et al., 2012b) and rain forests (Edwards et al., 2013; Nölscher et al., 2016) as well as in urban environments in wintertime (e.g. Dolgorouky et al., 2012; Yoshino et al., 2012). Results here show that measurements by the LIF instruments are accurate. In these comparison campaigns, deviations that are seen mostly for CRM instruments for complex conditions involving large concentrations of terpenic compounds show the tendency that OH reactivity is underestimated. Therefore, results do not indicate that high, unexplained reactivity values that were measured in previous field campaigns were due to measurement artifacts.

7 Data availability

Data are available on request from Hendrik Fuchs (h.fuchs@fz-juelich) and will be made available in the Eurochamp database (www.eurochamp.org).

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (SARLEP grant agreement No. 681529). This study was supported by the European Commission's FP7 programme (PIMMS grant agreement No. 287382). C. Schoemaecker, M. Blocquet, S. Coudert, S. Batut and C. Fittschen acknowledge financial support by the French ANR agency under contract No. ANR-11-LabEx-0005-01 CaPPA (Chemical and Physical Properties of the Atmosphere) and by the CNRS through PRC project No. 233277. S. Dusanter and N. Locoge acknowledge support from the French National Research Agency (ANR–11–LABX–0005–01) and the European Funds for Regional Economic Development (FEDER) through the CaPPA (Chemical and Physical Properties of the Atmosphere) project, and support from the Région Hauts-de-France and the Ministère de l'Enseignement Supérieur et de la Recherche (CPER Climibio) and the European Fund for Regional Economic Development. D. Cryer, C. Brumby, L. Whalley, D. Stone, P. Seakins and D. Heard acknowledge financial support from the Natural Environment Research Council (NERC) under grant numbers NE/J008990/1, NE/H003193/1 and NE/D004756/1, and also the NERC funded National Centre for Atmospheric Science W. Brune, A. Thames, and D. Miller acknowledge support from the US National Science Foundation grant AGS-1450838 and thank Hartwig Harder's group at Max-Planck-Institute for Chemistry in Mainz and Ronald Cohen's group at the University of California, Berkeley, for borrowing parts to complete the PSU instrument used in this comparison study. The authors thank Mathias Bachner (FZJ), Georg Stange (DWD) and Markus Rudolf (MPI) and Ronald Cohen's group (University of California, Berkeley) for technical support. The authors acknowledge the participation of Heidi Helen and Arnaud Praplan in the campaign in 2015.

References

Atkinson, R.: Gase-phase tropospheric chemistry of volatile organic compounds: 1. alkanes and alkenes, J. Phys. Chem. Ref. Data, 26, 217–290, 1997.

Atkinson, R., Aschmann, S. M., and Pitts, J. N.: Rate constants for the gas-phase reactions of the OH radical with a series of monoterpenes at 294±1 K, Int. J. Chem. Kin., 18, 287–299, doi:10.1002/kin.550180303, 1986.

Atkinson, R., Aschmann, S. M., and Arey, J.: Rate constants for the gas-phase reactions of OH and NO$_3$ radicals and O$_3$ with sabinene and camphene at 296±2 K, Atmos. Environ., 24, 2647–2654, doi:10.1016/0960-1686(90)90144-C, 1990.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., and Troe, J.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume I - gas phase reactions of O$_x$, HO$_x$, NO$_x$ and SO$_x$ species, Atmos. Chem. Phys., 4, 1461–1738, doi:10.5194/acp-4-1461-2004, 2004.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., Troe, J., and Subcommittee, I.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II - gas phase reactions of organic species, Atmos. Chem. Phys., 6, 3625–4055, doi:10.5194/acp-6-3625-2006, 2006.

Berresheim, H., Elste, T., Plass-Dülmer, C., Eisele, F. L., and Tanner, D. J.: Chemical ionization mass spectrometer for long-term measurements of atmospheric OH and H$_2$SO$_4$, Int. J. Mass. Spectrom., 202, 91–109, doi:10.1016/S1387-3806(00)00233-5, 2000.

Butkovskaya, N. I., Kukui, A., and Le Bras, G.: Branching fractions for H$_2$O forming channels of the reaction of OH radicals with acetaldehyde, J. Phys. Chem., 108, 1160–1168, doi:10.1021/jp036740m, 2004.

Cameron, M., Sivakumaran, V., Dillon, T. J., and Crowley, J. N.: Reaction between OH and CH$_3$CHO Part 1. Primary product yields of CH$_3$ (296 K), CH$_3$CO (296 K), and H (237-296 K), Phys. Chem. Chem. Phys., 4, 3628–3638, doi:10.1039/B202586H, 2002.

Chatani, S., Shimo, N., Matsunaga, S., Kajii, Y., Kato, S., Nakashima, Y., Miyazaki, K., Ishii, K., and Ueno, H.: Sensitivity analyses of OH missing sinks over Tokyo metropolitan area in the summer of 2007, Atmos. Chem. Phys., 9, 8975–8986, doi:10.5194/acp-9-8975-2009, 2009.

Di Carlo, P., Brune, W. H., Martinez, M., Harder, H., Lesher, R., Ren, X., Thornberry, T., Carroll, M. A., Young, V., Shepson, P. B., Riemer, D., Apel, E., and Campbell, C.: Missung OH reactivity in a forest: evidence for unkown reactiv biogenic VOCs, Science, 304, 722–726, doi:10.1126/science.1094392, 2004.

Dillon, T. J. and Crowley, J. N.: Direct detection of OH formation in the reactions of HO$_2$ with CH$_3$C(O)O$_2$ and other substituted peroxy radicals, Atmos. Chem. Phys., 8, 4877–4889, doi:10.5194/acp-8-4877-2008, 2008.

Dolgorouky, C., Gros, V., Sarda-Esteve, R., Sinha, V., Williams, J., Marchand, N., Sauvage, S., Poulain, L., Sciare, J., and Bonsang, B.: Total OH reactivity measurements in Paris during the 2010 MEGAPOLI winter campaign, Atmos. Chem. Phys., 12, 9593–9612, doi:10.5194/acp-12-9593-2012, 2012.

Dorn, H. P., Apodaca, R. L., Ball, S. M., Brauers, T., Brown, S. S., Crowley, J. N., Dubé, W. P., Fuchs, H., Häseler, R., Heitmann, U., Jones, R. L., Kiendler-Scharr, A., Labazan, I., Langridge, J. M., Meinen, J., Mentel, T. F., Platt, U., Pöhler, D., Rohrer, F., Ruth, A. A., Schlosser, E., Schuster, G., Shillings, A. J. L., Simpson, W. R., Thieser, J., Tillmann, R., Varma, R., Venables, D. S., and Wahner, A.: Intercomparison of NO$_3$ radical detection instruments in the atmosphere simulation chamber SAPHIR, Atmos. Meas. Tech., 6, 1111–1140, doi:10.5194/amt-6-1111-2013, 2013.

Edwards, P. M., Evans, M. J., Furneaux, K. L., Hopkins, J., Ingham, T., Jones, C., Lee, J. D., Lewis, A. C., Moller, S. J., Stone, D., Whalley, L. K., and Heard, D. E.: OH reactivity in a South East Asian tropical rainforest during the Oxidant and Particle Photochemical Processes (OP3) project, Atmos. Chem. Phys., 13, 9497–9514, doi:10.5194/acp-13-9497-2013, 2013.

Fuchs, H., Hofzumahaus, A., Rohrer, F., Bohn, B., Brauers, T., Dorn, H.-P., Häseler, R., Holland, F., Kaminski, M., Li, X., Lu, K., Nehr, S., Tillmann, R., Wegener, R., and Wahner, A.: Experimental evidence for efficient hydroxyl radical regeneration in isoprene oxidation, Nature Geosci., 6, 1023–1026, doi:10.1038/NGEO1964, 2013.

Fuchs, H., Tan, Z., Hofzumahaus, A., Broch, S., Dorn, H. P., Holland, F., Künstler, C., Gomm, S., Rohrer, F., Schrade, S., Tillmann, R., and Wahner, A.: Investigation of potential interferences in the detection of atmospheric RO_X radicals by laser-induced fluorescence under dark conditions, Atmos. Meas. Tech., 9, 1431–1447, doi:10.5194/amt-9-1431-2016, 2016.

Fuchs, H., Tan, Z., Lu, K., Bohn, B., Broch, S., Brown, S. S., Dong, H., Gomm, S., Häseler, R., He, L., Hofzumahaus, A., Holland, F., Li, X., Liu, Y., Lu, S., Min, K. E., Rohrer, F., Shao, M., Wang, B., Wang, M., Wu, Y., Zeng, L., Zhang, Y., Wahner, A., and Zhang, Y.: OH reactivity at a rural site (Wangdu) in the North China Plain: contributions from OH reactants and experimental OH budget, Atmos. Chem. Phys., 17, 645–661, doi:10.5194/acp-17-645-2017, 2017.

Gill, K. J. and Hites, R. A.: Rate constants for the gas-phase reactions of the hydroxyl radical with isoprene, α- and β-pinene, and limonene as a function of temperature, J. Phys. Chem. A, 106, 2538–2544, doi:10.1021/jp013532q, 2002.

Goldstein, A. H. and Galbally, I. E.: Known and unexplored organic constituents in the Earth's atmosphere, Environ. Sci. Technol., 41, 1514–1521, doi:10.1021/es072476p, 2007.

Hansen, R. F., Griffith, S. M., Dusanter, S., Rickly, P. S., Stevens, P. S., Bertman, S. B., Carroll, M. A., Erickson, M. H., Flynn, J. H., Grossberg, N., Jobson, B. T., Lefer, B. L., and Wallace, H. W.: Measurements of total hydroxyl radical reactivity during CABINEX 2009 - Part 1: field measurements, Atmos. Chem. Phys., 14, 2923–2937, doi:10.5194/acp-14-2923-2014, 2014.

Hansen, R. F., Blocquet, M., Schoemaecker, C., Léonardis, T., Locoge, N., Fittschen, C., Hanoune, B., Stevens, P. S., Sinha, V., and Dusanter, S.: Intercomparison of the comparative reactivity method (CRM) and pump-probe technique for measuring total OH reactivity in an urban environment, Atmos. Meas. Tech., 8, 4243–4264, doi:10.5194/amt-8-4243-2015, 2015.

Heard, D. E. and Pilling, M. J.: Measurement of OH and HO_2 in the troposphere, Chem. Rev., 103, 5163–5198, doi:10.1021/cr020522s, 2003.

Hofzumahaus, A., Rohrer, F., Lu, K., Bohn, B., Brauers, T., Chang, C.-C., Fuchs, H., Holland, F., Kita, K., Kondo, Y., Li, X., Lou, S., Shao, M., Zeng, L., Wahner, A., and Zhang, Y.: Amplified trace gas removal in the troposphere, Science, 324, 1702–1704, doi:10.1126/science.1164566, 2009.

Hohaus, T., Kuhn, U., Andres, S., Kaminski, M., Rohrer, F., Tillmann, R., Wahner, A., Wegener, R., Yu, Z., and Kiendler-Scharr, A.: A new plant chamber facility, PLUS, coupled to the atmosphere simulation chamber SAPHIR, Atmos. Meas. Tech., 9, 1247–1259, doi:10.5194/amt-9-1247-2016, 2016.

Ingham, T., Goddard, A., Whalley, L. K., Furneaux, K. L., Edwards, P. M., Seal, C. P., Self, D. E., Johnson, G. P., Read, K. A., Lee, J. D., and Heard, D. E.: A flow-tube based laser-induced fluorescence instrument to measure OH reactivity in the troposphere, Atmos. Meas. Tech., 2, 465–477, doi:10.5194/amt-2-465-2009, 2009.

IUPAC: IUPAC Task Group on Atmospheric Chemical Kinetic Data Evaluation, http://iupac.pole-ether.fr, 2017.

Kaiser, J., Skog, K. M., Baumann, K., Bertman, S. B., Brown, S. B., Brune, W. H., Crounse, J. D., de Gouw, J. A., Edgerton, E. S., Feiner, P. A., Goldstein, A. H., Koss, A., Misztal, P. K., Nguyen, T. B., Olson, K. F., St. Clair, J. M., Teng, A. P., Toma, S., Wennberg, P. O., Wild,

R. J., Zhang, L., and Keutsch, F. N.: Speciation of OH reactivity above the canopy of an isoprene-dominated forest, Atmos. Chem. Phys., 16, 9361–9379, doi:10.5194/acp-16-9361-2016, 2016.

Kim, S., Guenther, A., Karl, T., and Greenberg, J.: Contributions of primary and secondary biogenic VOC to total OH reactivity during the CABINEX (Community Atmosphere-Biosphere INteractions Experiments)-09 field campaign, Atmos. Chem. Phys., 11, 8613–8623, doi:10.5194/acp-11-8613-2011, 2011.

Kim, S., Sanchez, D., Wang, M., Seco, R., Jeong, D., Hughes, S., Barletta, B., Blake, D. R., Jung, J., Kim, D., Lee, G., Lee, M., Ahn, J., Lee, S.-D., Cho, G., Sung, M.-Y., Lee, Y.-H., Kim, D. B., Kim, Y., Woo, J.-H., Jo, D., Park, R., Park, J.-H., Hong, Y.-D., and Hong, J.-H.: OH reactivity in urban and suburban regions in Seoul, South Korea - an East Asian megacity in a rapid transition, Faraday Discuss., 189, 231–251, doi:10.1039/C5FD00230C, 2016.

Kovacs, T. A. and Brune, W. H.: Total OH loss rate measurement, J. Atmos. Chem., 39, 105–122, doi:10.1023/A:1010614113786, 2001.

Kumar, M., Busch, D. H., Subramaniam, B., and Thompson, W. H.: Criegee intermediate reaction with CO: Mechanism, barriers, conformer-dependence, and implications for ozonolysis chemistry, J. Chem. Phys. A, 118, 1887–1894, doi:10.1021/jp500258h, 2014.

Lou, S., Holland, F., Rohrer, F., Lu, K., Bohn, B., Brauers, T., Chang, C. C., Fuchs, H., Häseler, R., Kita, K., Kondo, Y., Li, X., Shao, M., Zeng, L., Wahner, A., Zhang, Y., Wang, W., and Hofzumahaus, A.: Atmospheric OH reactivities in the Pearl River Delta - China in summer 2006: measurement and model results, Atmos. Chem. Phys., 10, 11 243–11 260, doi:10.5194/acp-10-11243-2010, 2010.

Mao, J., Ren, X., Brune, W. H., Olson, J. R., Crawford, J. H., Fried, A., Huey, L. G., Cohen, R. C., Heikes, B., Singh, H. B., Blake, D. R., Sachse, G. W., Diskin, G. S., Hall, S. R., and Shetter, R. E.: Airborne measurement of OH reactivity during INTEX-B, Atmos. Chem. Phys., 9, 163–173, doi:10.5194/acp-9-163-2009, 2009.

Martinez, M., Harder, H., Kovacs, T. A., Simpas, J. B., Bassis, J., Lesher, R., Brune, W. H., Frost, G. J., Williams, E. J., Stroud, C. A., Jobson, B. T., Roberts, J. M., Hall, S. R., Shetter, R. E., Wert, B., Fried, A., Alicke, B., Stutz, J., Young, V. L., White, A. B., and Zamora, R. J.: OH and HO$_2$ concentrations, sources, and loss rates during the Southern Oxidants Study in Nashville, Tennessee, summer 1999, J. Geophys. Res., 108, 4617, doi:10.1029/2003JD003551, 2003.

MCM: Master Chemical Mechanism, http://mcm.leeds.ac.uk/MCM/, 2017.

Mehta, D., Nguyen, A., Montenegro, A., and Li, Z.: A kinetic study of the reaction of OH with xylenes using the relative rate/discharge flow/mass spectrometry technique, J. phys. Chem. A, 113, 12 942–12 951, doi:10.1021/jp905074j, 2009.

Michoud, V., Colomb, A., Borbon, A., Miet, K., Beekmann, M., Camredon, M., Aumont, B., Perrier, S., Zapf, P., Siour, G., Ait-Helal, W., Afif, C., Kukui, A., Furger, M., Dupont, J. C., Haeffelin, M., and Doussin, J. F.: Study of the unknown HONO daytime source at a European suburban site during the MEGAPOLI summer and winter field campaigns, Atmos. Chem. Phys., 14, 2805–2822, doi:10.5194/acp-14-2805-2014, 2014.

Michoud, V., Hansen, R. F., Locoge, N., Stevens, P. S., and Dusanter, S.: Detailed characterizations of the new Mines Douai comparative reactivity method instrument via laboratory experiments and modeling, Atmos. Meas. Tech., 8, 3537–3553, doi:10.5194/amt-8-3537-2015, 2015.

Muller, J. B. A., Kubistin, D., Elste, T., and Plaß Dülmer, C.: A novel semi-direct method to measure OH reactivity by chemical ionisation mass spectrometry (CIMS): in prep., Atmos. Meas. Tech. Discuss., 2017.

Nölscher, A. C., Sinha, V., Bockisch, S., Klüpfel, T., and Williams, J.: Total OH reactivity measurements using a new fast Gas Chromatographic Photo-Ionization Detector (GC-PID), Atmos. Meas. Tech., 5, 2981–2992, doi:10.5194/amt-5-2981-2012, 2012a.

Nölscher, A. C., Williams, J., Sinha, V., Custer, T., Song, W., Johnson, A. M., Axinte, R., Bozem, H., Fischer, H., Pouvesle, N., Phillips, G., Crowley, J. N., Rantala, P., Rinne, J., Kulmala, M., Gonzales, D., Valverde-Canossa, J., Vogel, A., Hoffmann, T., Ouwersloot, H. G., Vila-

Guerau de Arellano, J., and Lelieveld, J.: Summertime total OH reactivity measurements from boreal forest during HUMPPA-COPEC 2010, Atmos. Chem. Phys., 12, 8257–8270, doi:10.5194/acp-12-8257-2012, 2012b.

Nölscher, A. C., Bourtsoukidis, E., Bonn, B., Kesselmeier, J., Lelieveld, J., and Williams, J.: Seasonal measurements of total OH reactivity emission rates from Norway spruce in 2011, Biogeosciences, 10, 4241–4257, doi:10.5194/bg-10-4241-2013, 2013.

5 Nölscher, A. C., Butler, T., Auld, J., Veres, P., Munoz, A., Taraborrelli, D., Vereecken, L., Lelieveld, J., and Williams, J.: Using total OH reactivity to assess isoprene photooxidation via measurement and model, Atmos. Environ., 89, 453–463, doi:10.1016/j.atmosenv.2014.02.024, 2014.

Nölscher, A. C., Yanez-Serrano, A. M., Wolff, S., de Araujo, A. C., Lavric, J. V., Kesselmeier, J., and Williams, J.: Unexpected seasonality in quantity and composition of Amazon rainforest air reactivity, Nat. Commun., 7, 10 383, doi:10.1038/ncomms10383, 2016.

10 Onel, L., Blitz, M., Dryden, M., Thonger, L., and Seakins, P.: Branching ratios in reactions of OH radicals with methylamine, dimethylamine, and ethylamine, Environ Sci. Technol., 48, 9935–9942, doi:10.1021/es502398r, 2014.

Pang, X., Lewis, A. C., Rickard, A. R., Baeza-Romero, M. T., Adams, T. J., Ball, S. M., Daniels, M. J. S., Goodall, I. C. A., Monks, P. S., Peppe, S., Ródenas García, M., Sánchez, P., and Muñoz, A.: A smog chamber comparison of a microfluidic derivatisation measurement of gas-phase glyoxal and methylglyoxal with other analytical techniques, Atmos. Meas. Tech., 7, 373–389, doi:10.5194/amt-7-373-2014,
15 2014.

Parker, A. E., Amedro, D., Schoemaecker, C., and Fittschen, C.: OH radical reactivity measurements by FAGE, Environ. Engin. Manag. J., 10, 107–114, 2011.

Praplan, A. P., Pfannerstill, E., Williams, J., and Helen, H.: OH reactivity of the urban air in Helsinki, Finland, during winter: under review., Atmos. Environ., 2017.

20 Praske, E., Crounse, J. D., Bates, K. H., Kurten, T., Kjaergaard, H. G., and Wennberg, P. O.: Atmospheric fate of methyl vinyl ketone: Peroxy radical reactions with NO and HO_2, J. Phys. Chem. A, 119, 4562–4572, doi:10.1021/jp5107058, 2015.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipies in C, Cambridge University Press, 2nd Edition, 1992.

Ren, X., Harder, H., Martinez, M., Lesher, R. L., Oliger, A., Shirley, T., Adams, J., Simpas, J. B., and Brune, W. H.: HO_X concentra-
25 tions and OH reactivity observations in New York City during PMTACS-NY2001, Atmos. Environ., 37, 3627–3637, doi:10.1016/S1352-2310(03)00460-6, 2003.

Rohrer, F., Bohn, B., Brauers, T., Brüning, D., Johnen, F.-J., Wahner, A., and Kleffmann, J.: Characterisation of the photolytic HONO-source in the atmosphere simulation chamber SAPHIR, Atmos. Chem. Phys., 5, 2189–2201, doi:10.5194/acp-5-2189-2005, 2005.

Sadanaga, Y., Yoshino, A., Watanabe, K., Yoshioka, A., Wakazono, Y., Kanaya, Y., and Kajii, Y.: Development of a measurement
30 system of OH reactivity in the atmosphere by using a laser-induced pump and probe technique, Rev. Sci. Instr., 75, 2648–2655, doi:10.1063/1.1775311, 2004.

Schlosser, E., Brauers, T., Dorn, H.-P., Fuchs, H., Häseler, R., Hofzumahaus, A., Holland, F., Wahner, A., Kanaya, Y., Kajii, Y., Miyamoto, K., Nishida, S., Watanabe, K., Yoshino, A., Kubistin, D., Martinez, M., Rudolf, M., Harder, H., Berresheim, H., Elste, T., Plass-Dülmer, C., Stange, G., and Schurath, U.: Technical Note: Formal blind intercomparison of OH measurements: results from the international campaign
35 HOxComp, Atmos. Chem. Phys., 9, 7923–7948, doi:10.5194/acp-9-7923-2009, 2009.

Shirley, T. R., Brune, W. H., Ren, X., Mao, J., Lesher, R., Cardenas, B., Volkamer, R., Molina, L. T., Molina, M. J., Lamb, B., Velasco, E., Jobson, T., and Alexander, M.: Atmospheric oxidation in the Mexico City Metropolitan Area (MCMA) during April 2003, Atmos. Chem. Phys., 6, 2753–2765, doi:10.5194/acp-6-2753-2006, 2006.

Sinha, V., Williams, J., Crowley, J. N., and Lelieveld, J.: The comparative reactivity method – a new tool to measure total OH reactivity in ambient air, Atmos. Chem. Phys., 8, 2213–2227, doi:10.5194/acp-8-2213-2008, 2008.

Sinha, V., Williams, J., Diesch, J. M., Drewnick, F., Martinez, M., Harder, H., Regelin, E., Kubistin, D., Bozem, H., Hosaynali-Beygi, Z., Fischer, H., Andrés-Hernández, M. D., Kartal, D., Adame, J. A., and Lelieveld, J.: Constraints on instantaneous ozone production rates and regimes during DOMINO derived using in-situ OH reactivity measurements, Atmos. Chem. Phys., 12, 7269–7283, doi:10.5194/acp-12-7269-2012, 2012.

Stone, D., Whalley, L. K., Ingham, T., Edwards, P. M., Cryer, D. R., Brumby, C. A., Seakins, P. W., and Heard, D. E.: Measurement of OH reactivity by laser flash photolysis coupled with laser-induced fluorescence spectroscopy, Atmos. Meas. Tech., 9, 2827–2844, doi:10.5194/amt-9-2827-2016, 2016.

Whalley, L. K., Edwards, P. M., Furneaux, K. L., Goddard, A., Ingham, T., Evans, M. J., Stone, D., Hopkins, J. R., Jones, C. E., Karunaharan, A., Lee, J. D., Lewis, A. C., Monks, P. S., Moller, S. J., and Heard, D. E.: Quantifying the magnitude of a missing hydroxyl radical source in a tropical rainforest, Atmos. Chem. Phys., 11, 7223–7233, doi:10.5194/acp-11-7223-2011, 2011.

Whalley, L. K., Stone, D., Bandy, B., Dunmore, R., Hamilton, J. F., Hopkins, J., Lee, J. D., Lewis, A. C., and Heard, D. E.: Atmospheric OH reactivity in central London: observations, model predictions and estimates of in situ ozone production, Atmos. Chem. Phys., 16, 2109–2122, doi:10.5194/acp-16-2109-2016, 2016.

Williams, J. and Brune, W.: A roadmap for OH reactivity research, Atmos. Environ., 106, 371–372, doi:10.1016/j.atmosenv.2015.02.017, 2015.

Williams, J., Kessel, S. U., Nolscher, A. C., Yang, Y., Lee, Y., Yanez-Serrano, A. M., Wolff, S., Kesselmeier, J., Klupfel, T., Lelieveld, J., and Shao, M.: Opposite OH reactivity and ozone cycles in the Amazon rainforest and megacity Beijing: Subversion of biospheric oxidant control by anthropogenic emissions, Atmos. Environ., 125, Part A, 112–118, doi:10.1016/j.atmosenv.2015.11.007, 2016.

Winiberg, F. A. F., Dillon, T. J., Orr, S. C., Gro ß, C. B. M., Bejan, I., Brumby, C. A., Evans, M. J., Smith, S. C., Heard, D. E., and Seakins, P. W.: Direct measurements of OH and other product yields from the $HO_2 + CH_3C(O)O_2$ reaction, Atmos. Chem. Phys., 16, 4023–4042, doi:10.5194/acp-16-4023-2016, 2016.

Yang, Y., Shao, M., Wang, X., Nölscher, A. C., Kessel, S., Guenther, A., and Williams, J.: Towards a quantitative understanding of total OH reactivity: A review, Atmos. Environ., 134, 147–161, doi:10.1016/j.atmosenv.2016.03.010, 2016.

Yang, Y., Shao, M., Keßel, S., Li, Y., Lu, K., Lu, S., Williams, J., Zhang, Y., Zeng, L., Nölscher, A. C., Wu, Y., Wang, X., and Zheng, J.: How the OH reactivity affects the ozone production efficiency: case studies in Beijing and Heshan, China, Atmos. Chem. Phys., 17, 7127–7142, doi:10.5194/acp-17-7127-2017, 2017.

Yoshino, A., Nakashima, Y., Miyazaki, K., Kato, S., Suthawaree, J., Shimo, N., Matsunaga, S., Chatani, S., Apel, E., Greenberg, J., Guenther, A., Ueno, H., Sasaki, H., Hoshi, J.-y., Yokota, H., Ishii, K., and Kajii, Y.: Air quality diagnosis from comprehensive observations of total OH reactivity and reactive trace species in urban central Tokyo, Atmos. Environ., 49, 51–59, doi:10.1016/j.atmosenv.2011.12.029, 2012.

Zannoni, N., Dusanter, S., Gros, V., Sarda Esteve, R., Michoud, V., Sinha, V., Locoge, N., and Bonsang, B.: Intercomparison of two comparative reactivity method instruments inf the Mediterranean basin during summer 2013, Atmos. Meas. Tech., 8, 3851–3865, doi:10.5194/amt-8-3851-2015, 2015.

Zannoni, N., Gros, V., Lanza, M., Sarda, R., Bonsang, B., Kalogridis, C., Preunkert, S., Legrand, M., Jambert, C., Boissard, C., and Lathiere, J.: OH reactivity and concentrations of biogenic volatile organic compounds in a Mediterranean forest of downy oak trees, Atmos. Chem. Phys., 16, 1619–1636, doi:10.5194/acp-16-1619-2016, 2016.

Table 1. Specification of instrument parameters of OH reactivity instruments in these campaigns.

instrument	technique	time res. / s	1σ LOD[a] / s^{-1}	1σ accur.[a] (@ k_{OH})	flow rate / l/min	[OH] / 10^{10} cm^{-3}	inlet res. time/s	inlet line	reference
MDOUAI[b]	CRM[c]	600	1	18 %	0.37[d]	72[e,f]	5	1/4" PFA	Michoud et al. (2015)
LSCE[b]	CRM[c]	600	1	35 %	0.23[d]	78[e,f]	6	1/4" PFA	Zannoni et al. (2015)
MPI[b,g]	CRM[c]	900	1.6[b] / 1.3[g]	37 %	0.38	75[e,f]	6	1/4" PFA (heated)[b] / 1/2" PFA (heated)[g]	Sinha et al. (2008)
PSU[b]	FT-LIF[i]	30	0.5[j]	23 % ($< 2\,\mathrm{s}^{-1}$) / 8 % (10 to 100 s^{-1}) / 7 % ($> 100\,\mathrm{s}^{-1}$)	100	0.5[e]	1	1" PFA	Kovacs and Brune (2001) / Mao et al. (2009)
Lille[b]	LP-LIF[h]	30 to 60	0.4	8 %	9.5	0.6[k]	4	1/2" PFA	Parker et al. (2011)
Leeds[b]	LP-LIF[h]	100	0.4 to 1.0	6 %	16	3[k]	2.5	1/2" PFA	Stone et al. (2016)
FZJM[b]	LP-LIF[h]	40 to 160	0.2	10 %	15	0.7[k]	0.5	10 mm steel (Silconert coating)	Lou et al. (2010) / Fuchs et al. (2016)
FZJS[b,g]	LP-LIF[h]	60	0.1	10 %	20	0.8[k]	0.5	10 mm steel (Silconert coating)	Lou et al. (2010)
DWD[g]	FT-CIMS[l]	60 to 300	0.5	1 s^{-1} ($< 30\,\mathrm{s}^{-1}$) / 2 s^{-1} (30 to 40 s^{-1})	2280[m]	0.01[e]		no additional inlet line	Berresheim et al. (2000)

[a] limit of detection / accuracy as stated by the operator
[b] October 2015
[c] comparative reactivity measurement
[d] faster flow of 1 l/min in inlet line
[e] produced continuously by water photolysis (185 radiation of a pen-ray lamp)
[f] derived from the difference in the C1 and C2 measurement
[g] April 2016
[h] laser flash photolysis and laser-induced fluorescence
[i] flow-tube and laser-induced fluorescence
[j] limit of detection without the dilution, which amplifies this number by a factor of 5
[k] peak value produced by flash ozone photolysis (266 nm of a quadrupled Nd:YAG laser)
[l] flow-tube and chemical-ionization mass-spectrometry
[m] sampling rate from the chamber

Table 2. Specification of instruments measuring OH reactant concentrations in the two campaigns.

OH reactant	measurement	$1\,\sigma$ accuracy	$1\,\sigma$ precision	reaction rate constant
CO	Piccarro CRDS, GC (RGA)	15 ppbv, 8 %	5 ppbv	Atkinson et al. (2006); IUPAC (2017)
NO	chemiluminescence	5 %	4 pptv	Atkinson et al. (2004); IUPAC (2017)
NO_2	chemiluminescence	5 %	4 pptv	Atkinson et al. (2004); IUPAC (2017)
O_3	UV photometer	5 %	1 ppbv	Atkinson et al. (2006); IUPAC (2017)
CH_4	Piccarro CRDS	1 ppbv	1 ppbv	Atkinson et al. (2006); IUPAC (2017)
n-pentane	GC-FID	13 %	20 pptv	Atkinson et al. (2006); IUPAC (2017)
1-pentene	GC-FID, PTR-TOF-MS	13 %, 4 %	20 pptv, 19 pptv	Atkinson (1997); MCM (2017)
toluene	GC-FID, PTR-TOF-MS	13 %, 7 %	20 pptv, 7 pptv	Atkinson et al. (2006); IUPAC (2017)
o-xylene	GC-FID, PTR-TOF-MS	13 %, 2 %	10 pptv, 3 pptv	Mehta et al. (2009)
isoprene	GC-FID, PTR-TOF-MS	13 %, 6 %	20 pptv, 33 pptv	Atkinson et al. (2006); IUPAC (2017)
MVK	GC-FID	13 %	30 pptv	Atkinson et al. (2006); IUPAC (2017)
MACR	GC-FID	13 %	30 pptv	Atkinson et al. (2006); IUPAC (2017)
MVK+MACR	PTR-TOF-MS	6 %	22 pptv	
α-pinene	GC-FID	13 %	10 pptv	Atkinson et al. (2006); IUPAC (2017)
limonene	GC-FID	13 %	10 pptv	Gill and Hites (2002)
myrcene	GC-FID	13 %	10 pptv	Kim et al. (2011)
β-pinene	GC-FID	13 %	10 pptv	Gill and Hites (2002)
camphene	GC-FID	13 %	10 pptv	Atkinson et al. (1990)
Δ^3-carene	GC-FID	13 %	10 pptv	Atkinson et al. (1986)
β-ocimene	GC-FID	13 %	10 pptv	Kim et al. (2011)
β-phellandrene	GC-FID	13 %	10 pptv	Atkinson et al. (2006); IUPAC (2017)
sum monoterpenes	PTR-TOF-MS	4 %	5 pptv	
β-caryophyllene	GC-FID, PTR-TOF-MS	13 %, 6 %[a]	10 pptv, 15 pptv	Atkinson et al. (2006); IUPAC (2017)
HCHO	Hantzsch monitor	5 %	18 pptv	Atkinson et al. (2006); IUPAC (2017)
CH_3CHO	GC-FID, PTR-TOF-MS	13 %, 6 %[b]	200 pptv, 40 pptv	Atkinson et al. (2006); IUPAC (2017)

[a] PTR-TOF-MS measurements 40 % lower than GC-FID in 2016

[b] PTR-TOF-MS measurements 50 % higher than GC-FID in 2016

Table 3. Conditions during the experiments. Maximum concentrations during the experiments are given. Maximum values for OH reactivity are approximate values that do not refer to a specific instrument. Photolysis reactions were possible, if the chamber roof was open (mostly only part of the experiment). A mixture of aromatic compounds, alkenes and NO_x species is summarized as "urban" conditions.

date	added OH reactants	k_{OH}/s^{-1}	CO/ppbv	$\sum$VOC/ppbv	NO₂/ppbv	NO/ppbv	O₃/ppbv	$j_{NO_2}/10^{-3}s^{-1}$
05 October 2015	CO	200	33500	<LOD	<0.01	<0.1	85	dark
06 October 2015	CO, CH₄, NO₂	25	3300	11000	12	<0.03	85	dark
07 October 2015	CO, CH₄, NO	60	4100	13500	<1	120	<LOD	dark
09 October 2015	monoterpenes[a]	25	<LOD	11	<0.01	<0.05	295[b]	dark
11 October 2015	CO, isoprene	35	1650	9	0.06	0.7	50	3.0
	MVK[c], MACR[c]			2.3	3.2			
12 October 2015	urban[d], NO₂	40	<LOD	50	55	2.8	50	3.2
13 October 2015	CO, urban[d], NO₂	35	900	48	12.5	1.2	50	1.6
14 October 2015	monoterpenes[e]	15	<LOD	6	<0.08	<0.01	65[b]	dark
15 October 2015	CO, TME[f], NO	100	5400	40	15	30	45	dark
16 October 2015	β−caryophyllene,	30	<LOD	1.5	<0.04	<0.04	43[b]	dark
	CH₃CHO,			1.9				
	MVK, MACR			11				
07 April 2016	CO	60	16000	<LOD	<0.05	<0.01	80	dark
08 April 2016	n-pentane, NO	35	<LOD	60	<LOD	15	<LOD	dark
09 April 2016	isoprene, NO₂	45	<LOD	15	8	0.5	115	4.5
	MVK, MACR			7				
11 April 2016	urban[d], NO₂	50	<LOD	96	28	6	110	4.5
	CH₃CHO			45				
12 April 2016	urban[d], NO	60	<LOD	94	<LOD	32	<LOD	dark
	CH₃CHO			67				
13 April 2016	monoterpenes[a]	40	<LOD	19	<0.05	<0.02	185[b]	dark
14 April 2016	monoterpenes[a], NO₂	40	<LOD	16	29	2.8	70	4.3
15 April 2016	CO, β−caryophyllene	65	11500	4	<0.5	<0.04	45[b]	dark

[a] α-pinene, limonene, myrcene (liquid volume ratio 0.68:0.11:0.21)

[b] added at later times for ozonolysis

[c] photochemically formed

[d] o-xylene, toluene, 1-pentene (liquid volume ratio 1:1:1)

[e] identified compounds of plant emissions: α-pinene, β-pinene, limonene, myrcene, camphene, Δ^3-carene, ocimene, β-phellandrene

[f] 2,3-dimethyl-2-butene

Table 4. Correction applied to the raw data. Some corrections are non-linear and depend on several parameters (such as the pyrrole and OH concentrations). Values are given for typical atmospheric conditions. The instrument zero for the MDOUAI CRM was due to an contamination in the inlet system that was only present in this campaign. Corrections due to deviations from pseudo-first order reaction assumption depend on the actual OH reactivity value and specific VOC (see text for details). Typical numbers are given for 10 and $60\,\mathrm{s}^{-1}$. Interferences are present from NO, NO_2 and O_3 in some instruments. The correction depends on the concentrations of the interfering species often in a non-linear way. Therefore, only typical values can be given here.

instrument	instrument zero[a] / s^{-1}	humidity[a]	deviation from pseudo-first order[a] / s^{-1} (@ k_{OH})	10 ppbv NO[a] / s^{-1}	10 ppbv NO_2[a] / s^{-1}	dilution factor[b]	other
MDOUAI	-15[c]	$< 1\,\mathrm{s}^{-1}$	-0.5 (@ $10\,\mathrm{s}^{-1}$) -3 (@ $60\,\mathrm{s}^{-1}$)	10	3	1.32	
LSCE	0	$< 1\,\mathrm{s}^{-1}$	-1.6 (@ $10\,\mathrm{s}^{-1}$) -9.6 (@ $60\,\mathrm{s}^{-1}$)	17.5	1.9	1.45	
MPI	0	$< 1\,\mathrm{s}^{-1\,d}$	-2 to -2.5 (@ $10\,\mathrm{s}^{-1}$) -12 to -14 (@ $60\,\mathrm{s}^{-1}$)	20	1.6	1.3 to 1.9	O_3 interference 6 to $8\,\mathrm{s}^{-1\,a,e}$ (@ 50 ppbv O_3)
PSU	-3.4 ± 0.3	0	0	3^f	0	5^g	
Lille	-4.1 ± 0.4	0	0	0^h	0	$0, <1.7^i$	non-linearity -30 to -35 %[b,j]
Leeds	-2.3 ± 0.4	0	0	0	0	1.07^k	$\sim -2.5\,\mathrm{s}^{-1\,a,j}$
FZJM	-2.7 ± 0.2	0	0	0^h	0	$1.01^k, <3^i$	
FZJS	-1.3 ± 0.2	0	0	0^h	0	$1.01^k, <7^i$	
DWD	-8.3 ± 0.2	0	0	>8	0	0	NO contamination[l]

[a] absolute change

[b] relative change

[c] determined from periods of experiments without reactants, contamination in the inlet line only in this campaign

[d] up to $4\,\mathrm{s}^{-1}$ during fast humidity changes in 2016

[e] absolute change due to recycling of OH by ozone

[f] correction of the decay (Shirley et al., 2006)

[g] only applied in this campaign to reduce sampling flow rate from the chamber

[h] bi-exponential fit for NO> 20 ppbv (only applied on 07 October (Lille, FZJS, FZJM) and 15 October 2015 (FZJS, FZJM))

[i] only applied for dilution for high reactivity (Lille: $> 150\,\mathrm{s}^{-1}$, FZJS/FZJM: variable dilution for $> 60\,\mathrm{s}^{-1}$) on 05 October 2015

[j] deviations from single exponential OH decay likely due to misaligned photolysis beam in this campaign

[k] correction for added flow with O_3, when no O_3 was present in the experiment

[l] presence of 0.14 ppbv NO contamination within the instrument corrected for $k_{OH} > 2.5\,\mathrm{s}^{-1}$

Table 5. Fit results of the distribution of measurements to a Gaussian function when no OH reactants were present in the chamber. The distribution is either calculated by taking all data as they are measured or by forcing the mean values on each day to zero for an individual instrument. For the MDOUAI instrument, no independent instrument zero was determined.

instrument	data points	width (daily mean zero) / s^{-1}	width / s^{-1}	bias / s^{-1}
MDOUAI	46	2.3	n/a[a]	n/a[a]
LSCE	39	1.8	3.2	-0.4
MPI	105	2.1	4.6	1.8
PSU	798	0.9	1.1	1.1
Lille	758	0.2	0.5	0.5
Leeds	322	0.3	0.9	0.8
FZJM	585	0.1	0.2	0.2
FZJS	1450	0.2	0.3	0.0
DWD	230	0.3	0.4	-0.5

[a] not applicable, because measurements during zero air phases were used as instrument zero

Table 6. Results of the correlation analysis (linear correlation coefficient R^2 and slope and intercept of a weighted linear fit) for different subsets of the data. Errors of fit results (not shown here) are not significant within two digits of the fit parameters. $\langle|\Delta k|/k_{fit}\rangle$ gives the mean value of the relative difference between measurements and the regression line.

data subset	instrument	# data points	R^2	slope	intercept / s^{-1}	$\langle|\Delta k|/k_{fit}\rangle$
CO,CH$_4$[a]	MDOUAI	122	0.79	1.05	0.8	0.33
k(OH)< 60 s^{-1} [b]	LSCE	112	0.81	1.31	-0.1	0.54
	MPI	178	0.91	1.05	0.3	0.32
	PSU	1151	0.97	1.03	1.1	0.19[c]
	Lille	1506	0.97	1.04	1.3	0.10
	Leeds	829	0.98	1.06	1.3	0.11
	FZJM	795	0.98	1.12	0.2	0.10
	FZJS	2255	0.99	1.01	0.6	0.08
	DWD	295	0.99	0.87	-0.3	0.05
urban mix [d],	MDOUAI	145	0.85	0.99	0.8	0.17
isoprene [e,f]	LSCE	155	0.86	0.77	-0.1	0.17
	MPI	228	0.80	1.06	-1.1	0.17
	PSU	863	0.94	1.01	1.0	0.11[c]
	Lille	1355	0.97	0.94	0.4	0.08
	Leeds	677	0.98	1.07	0.1	0.09
	FZJM	993	0.98	0.94	0.1	0.06
	FZJS	reference				
	DWD	719	0.91	0.73	-0.3	0.16
monoterpenes	MDOUAI	69	0.48	0.56	2.4	0.45
no oxidation [f,g]	LSCE	69	0.68	0.82	-0.3	0.37
	MPI	73	0.72	0.70	1.5	0.34
	PSU	502	0.93	1.20	0.2	0.18[c]
	Lille	639	0.98	1.08	0.4	0.09
	Leeds	318	0.96	1.02	-0.1	0.16
	FZJM	585	0.99	0.96	0.2	0.07
	FZJS	reference				
	DWD	142	0.98	1.01	-0.6	0.09

[a] 05/06/07/11/13/15 October 2015; 07/15 April 2016

[b] reference: calculated reactivity

[c] scatter amplified by the dilution factor of 5 in this campaign

[d] 12/13 October 2015; 11 April 2016

[e] 11 October 2015; 09 April 2016

[f] reference: measurements by FZJS

[g] 09/14 October 2015; 13/14 April 2016
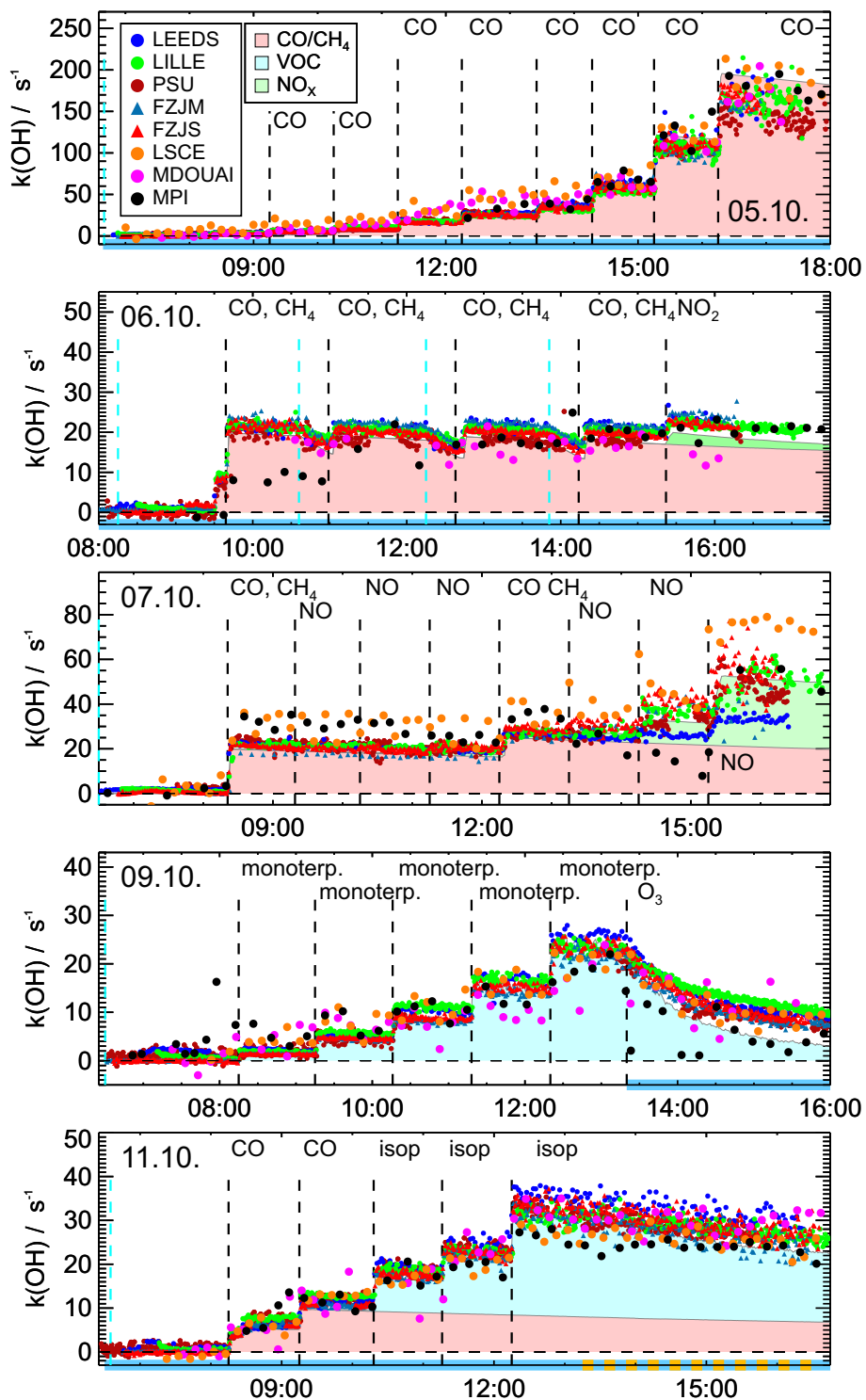
Figure 1. Measured (dots) and calculated (colored areas) OH reactivity during experiments in the SAPHIR chamber 2015. Error bars are omitted for the clarity of the plot, but are within the range of the scatter of data. Vertical dashed lines with labels give points in time of the injection of trace gases. Vertical dashed blue lines indicate addition of water, horizontal blue lines the presence of O_3 and horizontal dashed yellow lines illumination of the chamber by sunlight.
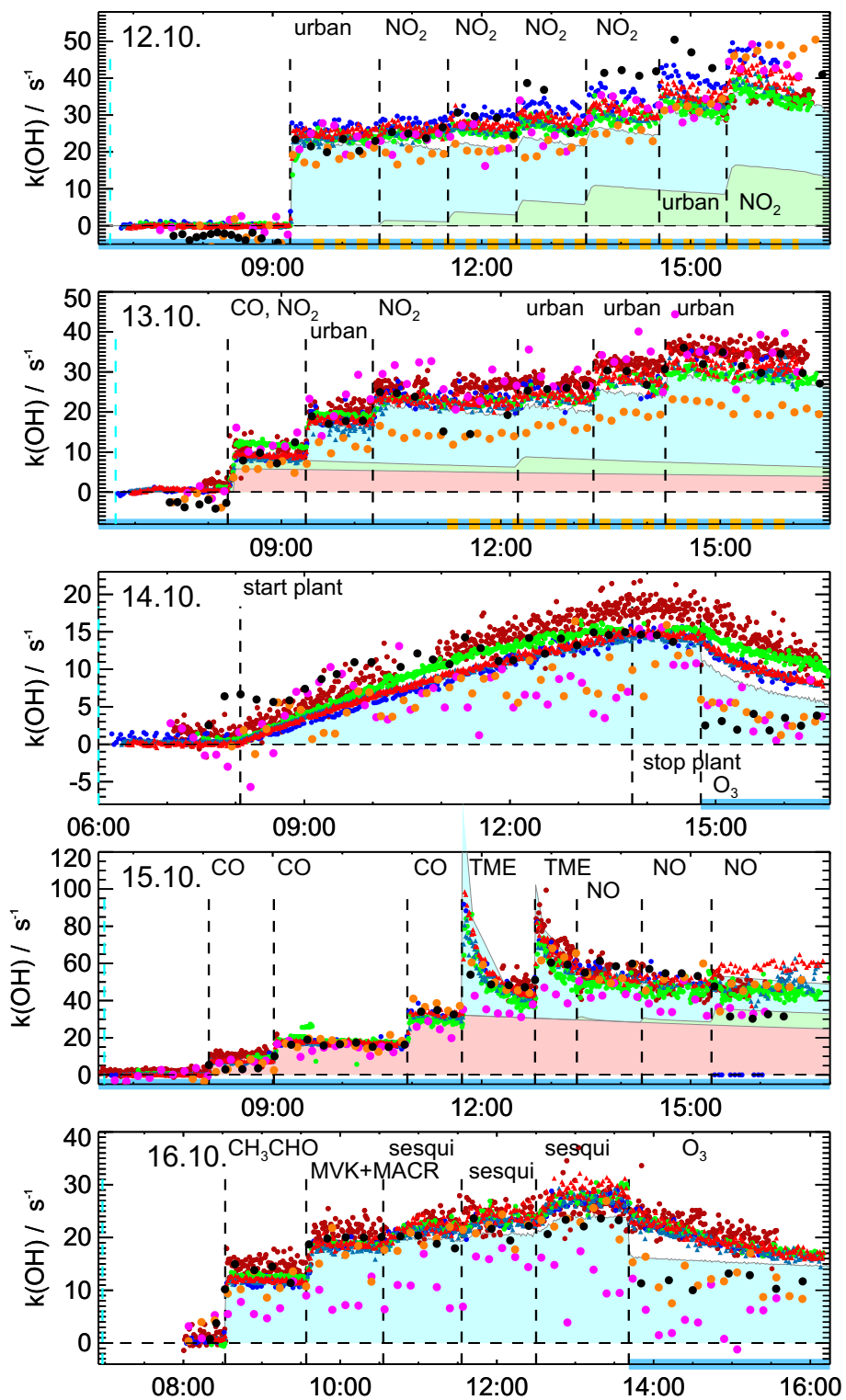
Figure 2. Same as Fig. 1. For the experiment on 16 October 2015 the subsequent addition of various VOC reactants, so that differences among measurements that only occur for a specific VOC reactant impact the visual agreement at later times, when other VOC reactants were injected.
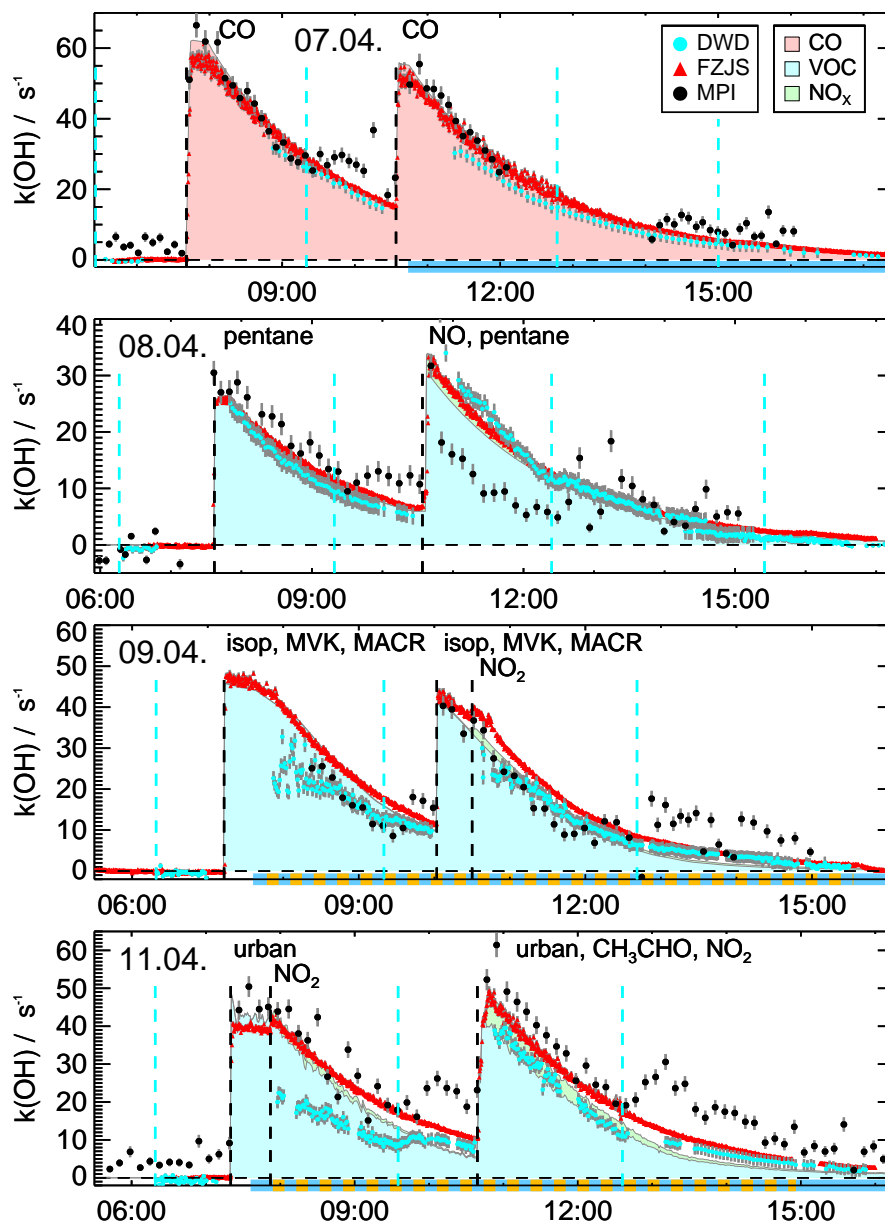
Figure 3. Measured (dots) and calculated (colored areas) OH reactivity during experiments in the SAPHIR chamber 2016. Vertical dashed lines with labels give points in time of the injection of trace gases. Vertical blue lines indicate addition of water, horizontal blue lines the presence of O_3 and horizontal yellow lines illumination of the chamber by sunlight.
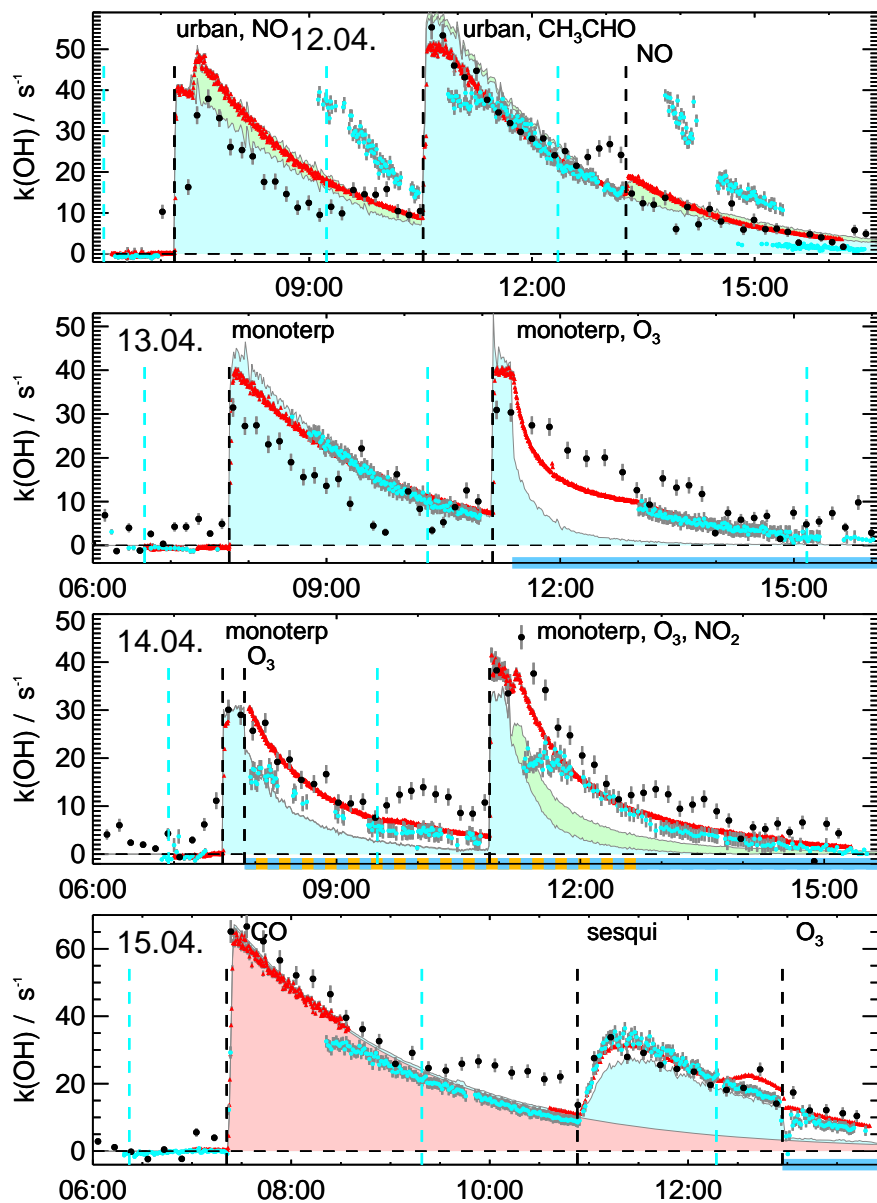
Figure 4. Measured (dots) and calculated (colored areas) OH reactivity during experiments in the SAPHIR chamber 2016. Vertical dashed lines with labels give points in time of the injection of trace gases. Vertical blue lines indicate addition of water, horizontal blue lines the presence of O_3 and horizontal yellow lines illumination of the chamber by sunlight.
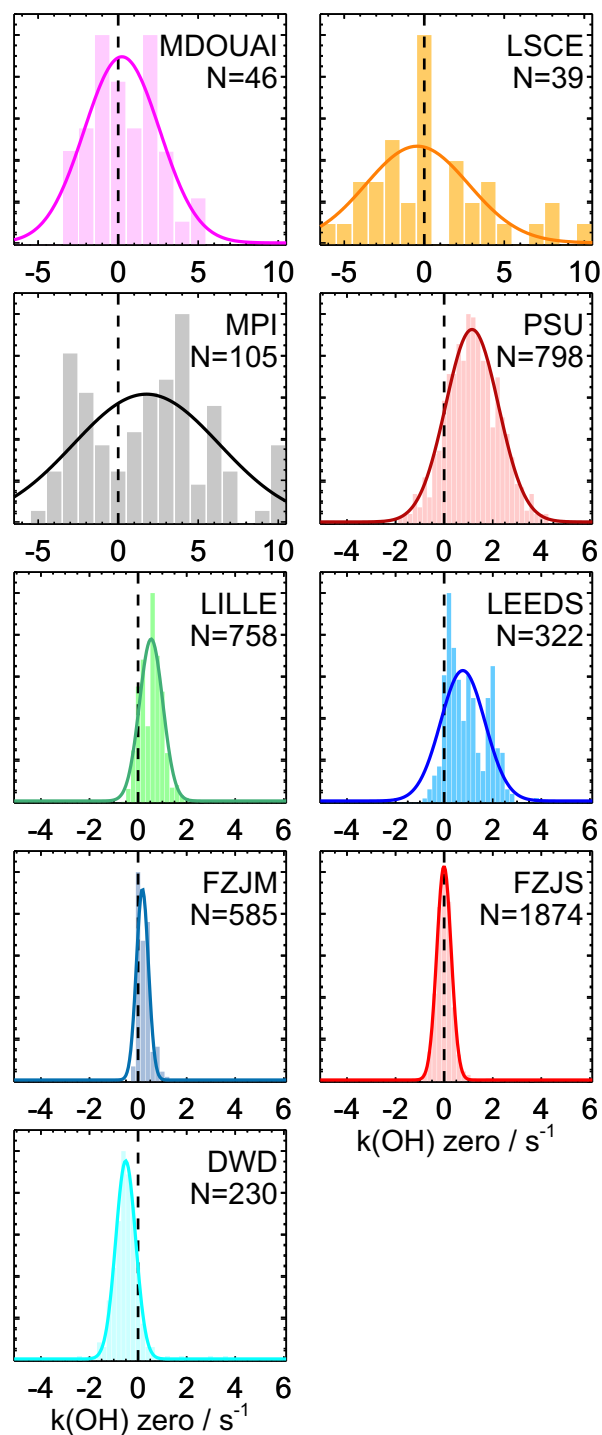
Figure 5. Distribution (bars) of OH reactivity measurements for times before OH reactants were injected into the chamber. MDOUAI, LSCE and MPI measurements are binned to $1\,\mathrm{s}^{-1}$ intervals, other measurements to $0.1\,\mathrm{s}^{-1}$ intervals. Solid lines shows the fit to a Gaussian distribution.
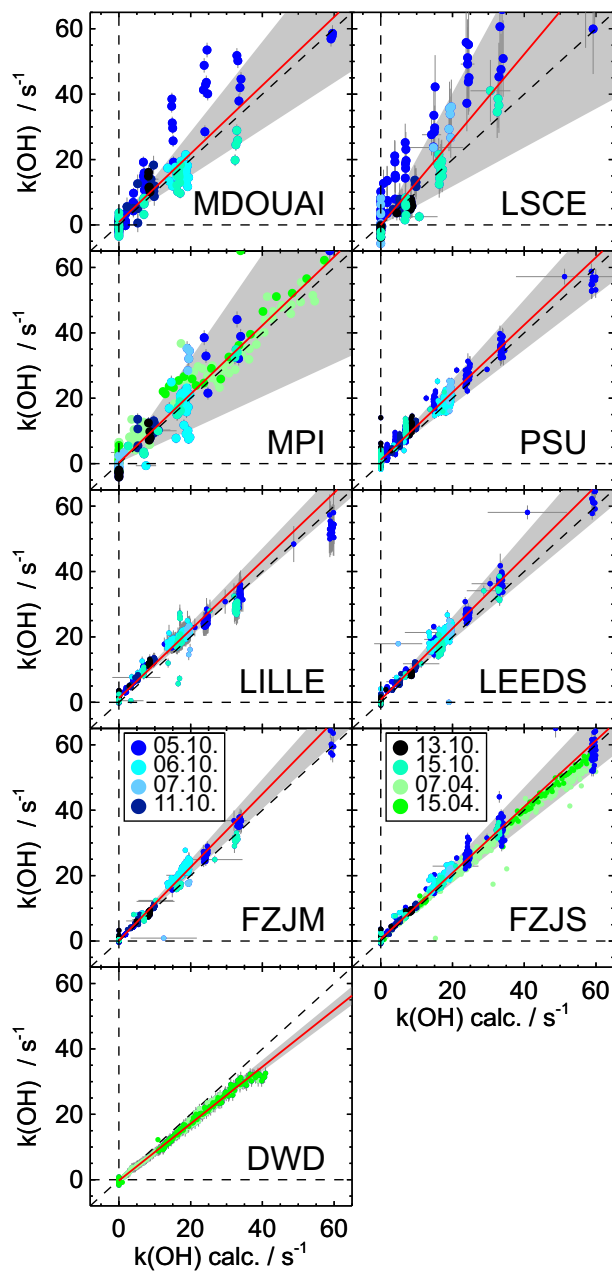
Figure 6. Correlation between measured and calculated OH reactivity for experiments in the SAPHIR chamber, if only CO, CH_4, O_3 and water vapor were present. Only the subset of the entire data set for values below $60\,s^{-1}$ is shown. Red lines give results of a linear regression analysis (Table 6). The grey area indicates the mean relative difference between measurements and the regression line. Measurements of the MDOUAI instrument during the first three experiments (05, 06 and 07 October 2015) have a higher uncertainty due to technical problems.
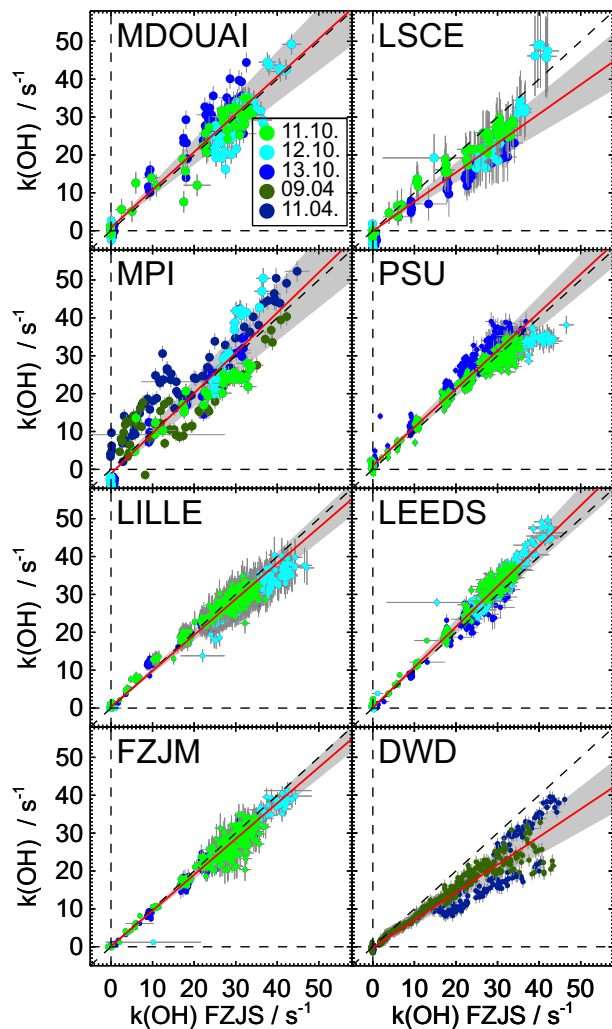
Figure 7. Correlation between measured OH reactivities and measurements by the FZJS instrument for experiments in the SAPHIR chamber when either isoprene (11 October 2015), isoprene, MVK and MACR (09 April 2016) or a mixture of propene, xylene and toluene (12/13 October 2015, 11 April 2016) were present in the chamber together with O_3 and water vapor. Lines give results of a linear regression analysis of the combined data set (Table 6). The grey area indicates the mean relative difference between measurements and the regression line.
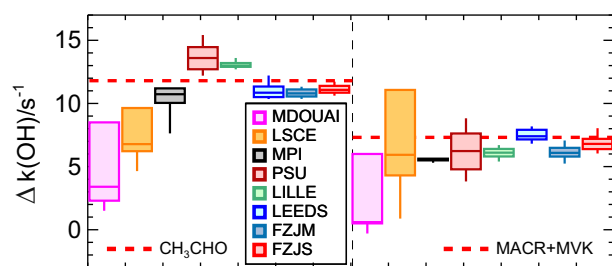
Figure 8. Change in reactivity measurements, when acetaldehyde and a mixture of MVK and MACR were added in the experiment on 15 October 2015. The box and whisker plot shows the statistics (median and 10, 25, 75, 90 percentiles) of 20 min time intervals 5 min before and after the injections of OH reactants. Dashed red lines give values calculated from the measured change of OH reactant concentrations.
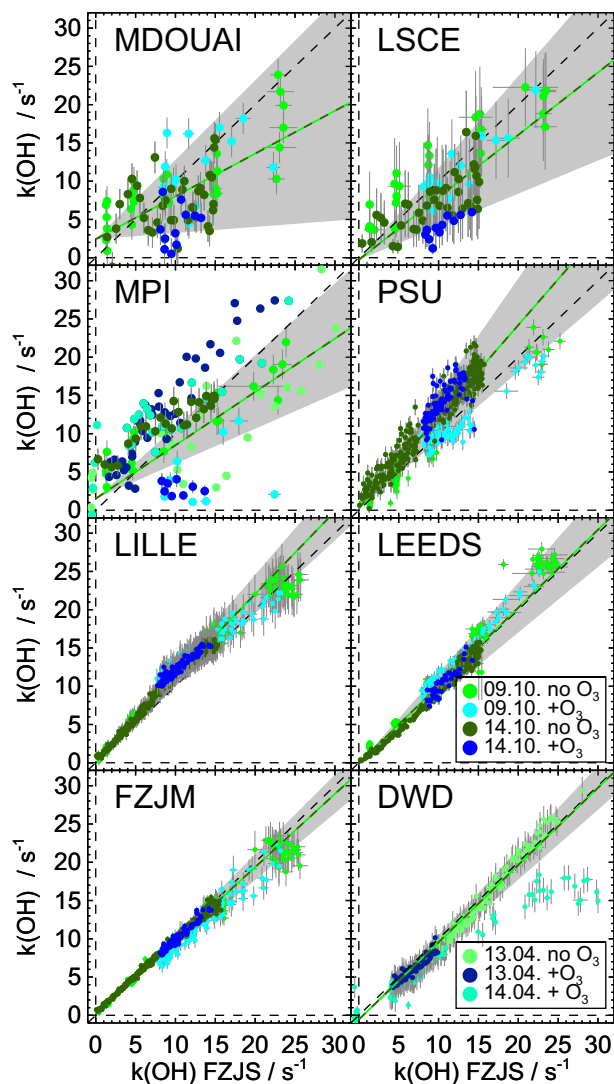
Figure 9. Correlation between measured OH reactivities and measurements by the FZJS instrument. This subset of data includes experiments with injections of a mix of monoterpenes ($\alpha-$pinene, limonene, and myrcene, 09 October 2015 and 13/14 April 2016) or with the transfer of real plant emissions, which consists predominantly of monoterpenes (14 October 2015). Times when O_3 was present are indicated by bluish colors. Lines give results of a linear regression analysis of a subset of data when no ozone was present (Table 6). The grey area indicates the mean relative difference between measurements and the regression line.
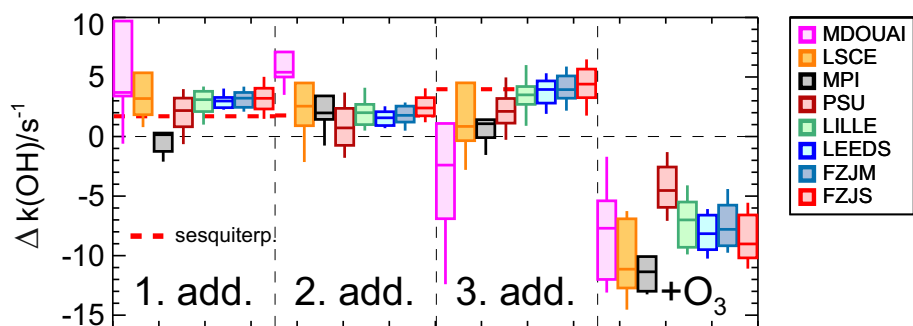
Figure 10. Change in reactivity measurements when $\beta-$caryophyllene was first added three times and of ozone in the last step during the experiment on 16 October 2015. The box and whisker plot shows the statistics (median and 10, 25, 75, 90 percentiles) of 20 min time intervals 5 min before and after each addition. Dashed red lines give values calculated from the measured change of OH reactant concentrations.
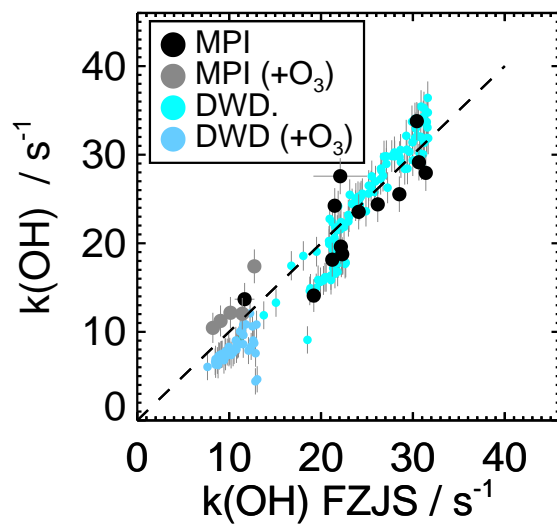
Figure 11. Correlation between OH reactivity measurements by the MPI CRM and DWD CIMS instruments and measurements by the FZJS instrument for the part of the experiment on 15 April 2016 when $\beta-$caryophyllene was present in the chamber. Times when ozone was also present are colored differently.
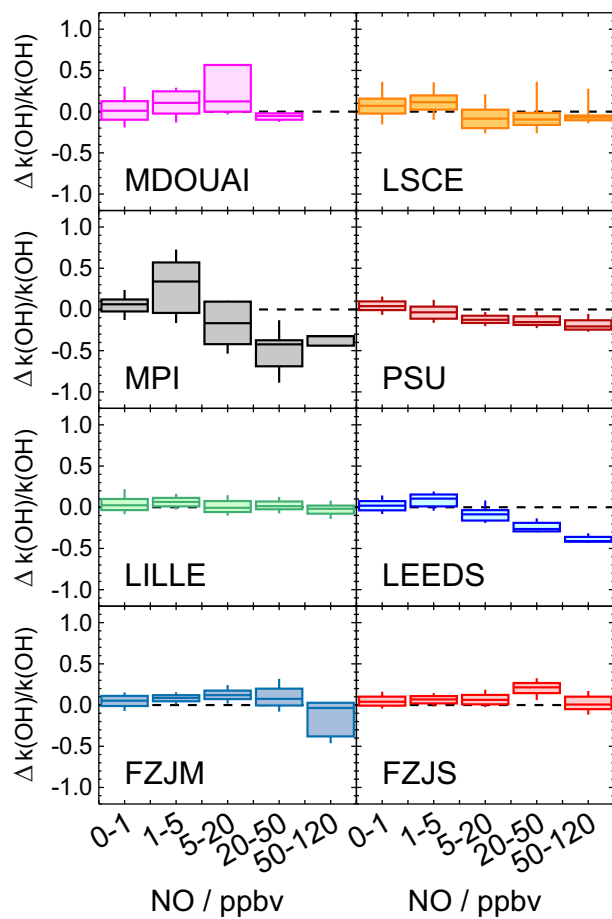
Figure 12. Relative difference between measured and calculated OH reactivity depending on the NO mixing ratio in experiments on 07, 12, 13, 15 October 2015. Boxes give median and 25 and 75 percentiles, whiskers give 10 and 90 percentiles. Because the MDOUAI CRM instrument did not measure on 07 October 2015, when NO concentrations were systematically changed, limited data are available for this instrument. For the Leeds LP-LIF instrument the OH decays were fitted by a single exponential decay for all NO, whereas a bi-exponential decay was used for Lille, FZJM and FZJS for NO > 20 ppbv.
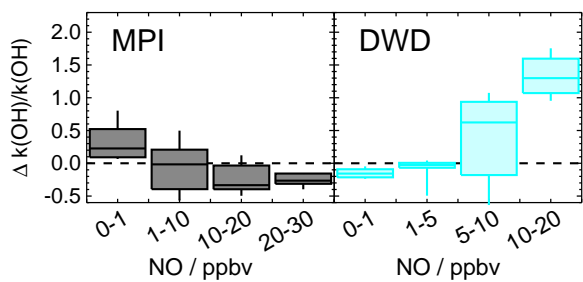
Figure 13. Relative difference between measured OH reactivity and measurements by the FZJS instrument depending on the NO mixing ratio for the experiment on 08/12 April 2016. Boxes give median and 25 and 75 percentiles, whiskers give 10 and 90 percentiles.
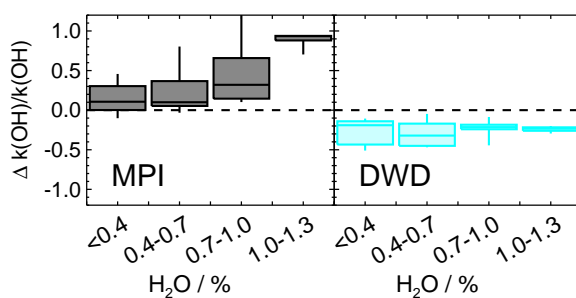
Figure 14. Relative difference between measured OH reactivity and measurements by the FZJS instrument depending on water vapor mixing ratio during experiments in April 2016 when only CO, pentane, or a mixture of propene, xylene and toluene were injected (part of the experiments on 07, 08, 11, 15 April 2016).