Mapping gas-phase organic reactivity and concomitant secondary organic aerosol formation: chemometric dimension reduction techniques for the deconvolution of complex atmospheric datasets

Kevin P. Wyche[1, 2*], Paul S. Monks[2], Kirsty L. Smallbone[1], Jacqueline F. Hamilton[3], M. Rami Alfarra[4,5], Andrew R. Rickard[3,6], Gordon B. McFiggans[4], Michael E. Jenkin[7], William J. Bloss[8], Annette C. Ryan[9], C. Nicholas Hewitt[9], A. Robert MacKenzie[10]

1. Air Environment Research Group, School of Environment and Technology, University of Brighton, BN2 4GJ, UK
2. Department of Chemistry, University of Leicester, Leicester, LE1 7RH, UK
3. Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York, YO10 5DD, UK
4. School of Earth, Atmospheric and Environmental Sciences, University of Manchester, M13 9PL, UK
5. National Centre for Atmospheric Science, University of Manchester, M13 9PL, UK
6. National Centre for Atmospheric Science, University of York, York, YO10 5DD, UK
7. Atmospheric Chemistry Services, Okehampton, Devon, EX20 1FB, UK
8. School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK
9. Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK
10. Birmingham Institute of Forest Research, University of Birmingham, B15 2TT, UK

*For correspondence:
K. P. Wyche
Air Environment Research Group, University of Brighton, Brighton, BN2 4GJ, UK.
Tel: +44 (0)1273 643306
Email: k.p.wyche@brighton.ac.uk

Abstract

Highly non-linear dynamical systems, such as those found in atmospheric chemistry, necessitate hierarchical approaches to both experiment and modeling in order, ultimately, to identify and achieve fundamental process-understanding in the full open system. Atmospheric simulation chambers comprise an intermediate in complexity, between a classical laboratory experiment and the full, ambient system. As such, they can generate large volumes of difficult-to-interpret data. Here we describe and implement a chemometric dimension reduction methodology for the deconvolution and interpretation of complex gas- and particle-phase composition spectra. The methodology comprises principal component analysis (PCA), hierarchical cluster analysis (HCA) and positive least squares-discriminant analysis (PLS-DA). These methods are, for the first time, applied to simultaneous gas- and particle-phase composition data obtained from a comprehensive series of environmental simulation chamber experiments focused on biogenic volatile organic compound (BVOC) photooxidation and associated secondary organic aerosol (SOA) formation. We primarily investigated the biogenic SOA precursors isoprene, α-pinene, limonene, myrcene, linalool and β-caryophyllene. The chemometric analysis is used to classify the oxidation systems and resultant SOA according to the controlling chemistry and the products formed. Results show that "model" biogenic oxidative systems can be successfully separated and classified according to their gaseous oxidation products. Furthermore, a holistic view of results across both the gas- and particle-phases shows the different SOA formation chemistry, initiating in the gas-phase, proceeding to govern the differences between the various BVOC SOA compositions. The results obtained are used to describe the particle composition in the context of the oxidized gas-phase matrix. An extension of the technique, which incorporates into the statistical models data from anthropogenic (*i.e.* toluene) oxidation and "more realistic" plant mesocosm systems, demonstrates that such an ensemble of chemometric mapping has the potential to be used for the classification of more complex spectra of unknown origin. More specifically, the addition of mesocosm data from fig and birch tree experiments shows that isoprene and monoterpene emitting sources, respectively, can be mapped onto the statistical model structure and their positional vectors can provide insight into their biological sources and controlling oxidative chemistry. The potential to extend the methodology to the analysis of ambient air is discussed using results obtained from a zero-dimensional box model incorporating mechanistic data obtained from the Master Chemical Mechanism (MCMv3.2). Such an extension to analysing ambient air would prove a powerful asset in assisting with the identification of SOA sources and the elucidation of the underlying chemical mechanisms involved.

Keywords: Volatile organic compounds, secondary organic aerosol, environmental simulation chamber, photooxidation, principal component analysis, cluster analysis, positive least-squares discriminant analysis, chemometrics, terpenes, mesocosm

1. Introduction

Biogenic Volatile Organic Compounds (BVOCs) are ubiquitous in the global troposphere, being emitted primarily from terrestrial plant life (Kanakidou et al., 2005). It is estimated that the total annual emission rate of all (non-methane) BVOCs is roughly ten times that of all anthropogenic volatile organic compounds, being around 750 Tg C yr^{-1} (Sindelarova et al., 2014). With the exception of methane, the most dominant species of BVOCs in terms of emission strength, reactivity and their impact upon the atmosphere, are terpenes (Reinnig et al., 2008) a subdivision of BVOCs that primarily comprise the hemiterpene, isoprene (C_5), monoterpenes (C_{10}) and sesquiterpenes (C_{15}) (*e.g.* (Atkinson and Arey, 2003a;Kanakidou et al., 2005)).

Within the troposphere terpenes are able to react with OH, O_3 and NO_3 at appreciable rates (*e.g.* (Calvert et al., 2000;Koch et al., 2000;Fantechi et al., 2002;Capouet et al., 2004;Kroll et al., 2006) such that their atmospheric lifetimes are in the order of minutes – hours (*e.g.* (Calogirou et al., 1999)). Because of their large emission rates and high reactivities, terpenes have a strong impact upon the chemistry of the troposphere at the local, regional and global scales (*e.g.* (Jaoui and Kamens, 2001;Paulot et al., 2012;Surratt, 2013). For instance, terpenes have high photochemical ozone creation potentials (Derwent et al., 2007) and extensive photochemical oxidation pathways that lead to the production of a complex array of oxygenated and nitrated products, some of which are able to form secondary organic aerosol (SOA) (*e.g.* (Calvert et al., 2000;Capouet et al., 2004;Jenkin, 2004;Baltensperger et al., 2008;Kanakidou et al., 2005;Surratt et al., 2006;Kroll and Seinfeld, 2008;Hallquist et al., 2009).

Aerosol particles are natural components of the Earth's atmosphere responsible for a range of well-documented impacts, ranging from visibility impairment on the local scale to climate change, with suspended particles being able to perturb the Earth's radiative budget *via* both direct and indirect mechanisms (IPCC, 2007). Furthermore, fine airborne particles have been shown to have numerous detrimental effects on human health, particularly in vulnerable members of the population (Harrison et al., 2010;Heal et al., 2012).

Biogenic SOA (BSOA) has been estimated to account for a significant fraction of total global SOA. Modelling studies suggest the annual global production rate of BSOA is of the order 16.4 Tg Yr^{-1} (Henze and Seinfeld, 2006). However, despite its importance and the significant amount of investigation conducted upon it, the formation mechanisms and chemical composition of BSOA are still not well characterised (*e.g.* (Librando and Tringali, 2005;Wang et al., 2013). Indeed under certain conditions as much as 80 – 90 % of analysed SOA mass is unknown (Limbeck et al., 2003;Kalberer et al., 2006). In particular, there remains a significant lack of information regarding the composition and evolution of the complex organic gas-phase matrix during aerosol formation, and its linkage to SOA (Kroll et al., 2005;Librando and Tringali, 2005). Indeed, in the many studies conducted on BSOA, very few oxidation products of the precursor are routinely identified and reported.

The chemistry of the atmospheric system is highly non-linear and can be studied by experiments ranging from highly controlled laboratory studies of a single process, to field studies of the whole complex system. A significant proportion of the findings gained regarding SOA over the last decade and more have come from atmospheric simulation chamber experiments, intermediate in complexity between classical single-process experiments and the fully open system (for various different chamber systems and VOC precursors, see for example, Pandis et al., 1991; Odum et al., 1996; Hoffmann et al., 1997; Griffin et al., 1999; Glasius et al., 2000; Cocker et al., 2001; Jaoui and Kamens, 2003; Kleindienst et al., 2004; Presto et al., 2005; Bloss et al., 2005; Rohrer et al., 2005; Ng et al., 2006, 2007; Dommen et al., 2006; Surrat et al., 2006; Grieshop et al., 2007; Chan et al., 2007; Wyche et al., 2009; Hildebrandt et al., 2009; Rickard et al., 2010; Camredon et al., 2010; Chhabra et al., 2011; Hennigan et al., 2011; Jenkin et al., 2012). Chamber experiments produce a large amount of data, the interpretation of which can often be highly complex and time consuming even though the set-up of the chamber constrains the complexity to a large degree.

In the current "big data" age, advanced monitoring techniques are producing increasingly larger, more complex and detailed data sets. Modern chamber experiments, monitored by state-of-the-art gas- and particle-phase instrumentation, often yield so much data that often only a fraction is subsequently used in a given analysis. For example, during a typical six-hour environmental simulation chamber experiment, VOC monitoring chemical ionisation reaction time-of-flight mass spectrometry (CIR-TOF-MS) will produce roughly 1.1 x 10^7 data points. In order to keep pace with instrument development and maximise the information extracted from sometimes-complex experiments, it is crucial that we advance our data analysis methods and introduce new data mining techniques.

The work reported here focuses on detailed organic gas-phase and particle-phase composition data, recorded during SOA atmospheric simulation chamber experiments, using CIR-TOF-MS and liquid chromatography-ion trap mass spectrometry (LC-MS/MS), respectively, as well as broad (*i.e.* generic composition "type"; oxygenated organic aerosol, nitrated, sulphated *etc*) aerosol composition data, recorded by compact time-of-flight aerosol mass spectrometry (cTOF-AMS). The goal of this paper is to demonstrate and evaluate the application of an ensemble reductive chemometric methodology for these comprehensive oxidation chamber datasets, to be used as a model framework to map chemical reactivity from mesocosm systems, thus providing a link from model systems to more "real" mixtures of organics. The intermediate complexity offered by simulation chamber experiments makes them an ideal test-bed for the methodology. Application of the methodology to resultant particle-phase data also aims to provide a level of particle composition classification in the context of gas-phase oxidation.

Similar approaches using statistical analyses have been recently applied to both detailed and broad ambient aerosol composition data (e.g. (Heringa et al., 2012;Paglione et al., 2014)), particularly in the context of source apportionment (e.g. (Alier et al., 2013)). Different methods have been attempted by several groups to deconvolve organic aerosol spectra measured by the Aerosol Mass Spectrometer (AMS) in particular (e.g. Zhang et al., 2005, 2007; Marcolli et al., 2006; Lanz et al., 2007). Zhang et al. (2005) applied a custom principal component analysis (CPCA) method to extract two distinct sources of organic aerosols in an urban environment using linear decomposition of AMS spectra and later applied a Multiple Component Analysis technique (MCA, an expanded version of the CPCA) to separate more than two factors in datasets from 37 field campaigns in the Northern Hemisphere (Zhang et al., 2007). Marcolli et al. (2006) applied a hierarchical cluster analysis method to an ambient AMS data set, and reported clusters representing biogenic VOC oxidation products, highly oxidised organic aerosols and other small categories. Receptor modelling techniques such as Positive Matrix Factorization (PMF) employ similar multivariate statistical methods in order to deconvolve a time series of simultaneous measurements into a set of factors and their time-dependent concentrations (Paatero and Tapper, 1994; Paatero, 1997). Depending on their specific chemical and temporal characteristics, these factors may then be related to emission sources, chemical composition and atmospheric processing. For example, Lanz et al. (2007) and Ulbrich et al., (2009) applied PMF to the organic fraction of AMS datasets and were able to conduct source apportionment analysis identifying factors contributing to the composition of organic aerosol at urban locations. Slowik et al. (2010), combined both particle-phase AMS and gas-phase proton transfer reaction mass spectrometry data for the PMF analysis of urban air, and were able to successfully obtain “regional transport, local traffic, charbroiling and oxidative process” factors. By combining the two datasets, Slowik and colleagues were able to acquire more in-depth information regarding the urban atmosphere than could be derived from the analysis of each of the sets of measurements on their own.

Because receptor models require no a priori knowledge of meteorological conditions or emission inventories, they are ideal for use in locations where emission inventories are poorly characterised or highly complicated (e.g. urban areas), or where atmospheric processing plays a major role. However, because all of the values in the profiles and contributions are constrained to be positive, the PMF model can have an arbitrary number of factors and the user must select the “best” solution that explains the data. This subjective step of PMF analysis relies greatly on the judgment and skill of the user.

The central methodology employed, is based around the application of *principal component analysis* (PCA), *hierarchical cluster analysis* (HCA) and *positive least squares-discriminate analysis* (PLS-DA) of single-precursor oxidant chemistry in environmental simulation chambers. Colloquially, we can describe these three approaches as providing dimensions along which the data are separable (PCA), tests of relatedness (HCA) and checks for false-positives (PLS-DA).

Such dimension reduction techniques can be very powerful when used in chemometrics, enabling large and often complex datasets to be rendered down to a relatively small set of pattern-vectors provide an optimal description of the variance of the data (Jackson, 1980;Sousa et al., 2013;Kuppusami et al., 2014). Unlike other statistical techniques such as PMF, the ensemble methodology presented here does not require the use of additional external databases (comprising information regarding different environments/reference spectra), is simpler to use and less labour intensive, and places less importance on user skill in the production of accurate and meaningful results. Moreover, the primary focus of techniques such as PMF is on source identification/separation, whereas here the focus is placed on compositional isolation.

The analysis conducted in this work shows that “model” biogenic oxidative systems can be clearly separated and classified according to their gaseous oxidation products, *i.e.* isoprene from β-caryophyllene from non-cyclic monoterpenes and cyclic monoterpenes. The addition of equivalent mesocosm data from fig and birch tree experiments shows that large isoprene and large monoterpene emitting sources, respectively, can be mapped onto the statistical model structure and their positional vectors can provide insight into the oxidative chemistry at play. The analysis is extended to particle-phase data to show further classifications of model systems based on both broad and detailed SOA composition measurements.

The methodology described and the results presented (supported by findings obtained from zero-dimensional box modelling), indicate that there is some potential that the approach could ultimately provide the foundations for a framework onto which it would be possible to map the chemistry and oxidation characteristics of ambient air measurements. This could in turn allow "pattern" typing and source origination for certain complex air matrices and provide a snapshot of the reactive chemistry at work, lending insight into the type of chemistry driving the compositional change of the contemporary atmosphere. There are similarities between this approach to discovery science in the atmosphere and metabolomics strategies in biology (*e.g.* (Sousa et al., 2013;Kuppusami et al., 2014)).

2. Experimental details

2.1 Choice of precursors

Six different BVOCs and one anthropogenic VOC were chosen for analysis. The target compounds, their structures and reaction rate constants with respect to OH and O_3 are given in Table 1. The BVOCs were chosen according to their atmospheric prevalence, structure and contrasting photooxidative reaction pathways; all have previously been shown to form SOA under simulation chamber conditions (e.g. Hoffmann et al., 1997; Griffin et al., 1999; Glasius et al., 2000; Jaoui and Kamens, 2003; Presto et al., 2005; Ng et al., 2006; Surratt et al., 2006; Dommen et al., 2006; Lee et al., 2006; Hallquist et al., 2009; Alfarra et al., 2013, and references therein). Isoprene is a C_5 diene that accounts for around 62 % (~ 594 Tg yr^{-1}) of total annual non-methane BVOC emissions (Sindelarova et al., 2014). After isoprene, monoterpenes (C_5H_{16}) have the next largest annual emission rate, they account for around 11 % (~ 95 Tg yr^{-1}) of total annual non-methane BVOC emissions (Sindelarova et al., 2014). α-pinene and limonene were chosen for analysis here alongside isoprene, the former acting as a model system to represent bicyclic monoterpenes, the later to represent monocyclic diene terpenes. In this work, α-pinene and limonene together generically represent (and are referred to hereafter as) "cyclic" monoterpenes (*i.e.* monoterpenes that contain one six-member carbon ring). In order to explore the chemistry of non-cyclic monoterpenes, myrcene, an acyclic triene monoterpene, was also included, as was the structurally similar acyclic diene

OVOC, linalool. In this work, myrcene and linalool together generically represent (and are referred to hereafter as) “straight chain” monoterpenes/BVOCs (note: linalool is not technically a monoterpene, but does contain the same carbon backbone as myrcene, consequently it is expected to exhibit similar photooxidative chemistry). Finally, β-caryophyllene was included to represent sesquiterpenes, which have annual emissions of the order 20 Tg yr^{-1} (Sindelarova et al., 2014). In order to test the ability of the methodology to distinguish between biogenic and anthropogenic systems, toluene was also included. Toluene is often used as a model system to act as a proxy for aromatic species in general (Bloss et al., 2005). For contrasting plant mesocosm systems, *Ficus benjamina* and *Ficus cyathistipula* (fig) and *Betula pendula* (birch) species were chosen to represent tropical rainforest and European environs, respectively.

In general, the VOC precursors employed have roughly similar reaction rate constants with respect to OH and O_3, *e.g.* limonene, myrcene, linalool and β-caryophyllene all have atmospheric lifetimes with respect to OH of the order 40 – 50 minutes (Alfarra et al., 2013;Atkinson and Arey, 2003b). β-caryophyllene has the shortest lifetime with respect to O_3 (*ca.* 2 minutes) and isoprene and α-pinene have the longest lifetimes with respect to both OH and O_3, *e.g.* isoprene and α-pinene have atmospheric lifetimes with respect to OH of the order 1.4 – 2.7 hours (Alfarra et al., 2013;Atkinson and Arey, 2003b). In order to ensure the various systems had progressed sufficiently down their respective photooxidative reaction pathways, the experiment duration was set to be sufficiently long that the majority of the precursor had been consumed by the conclusion of the experiment.

2.2 Chamber Infrastructure

Experiments were carried out across three different European environmental simulation chamber facilities over a number of separate campaigns. The chambers used, included (1) The University of Manchester Aerosol Chamber (MAC), UK (Alfarra et al., 2012); (2) The European Photoreactor (EUPHORE), ES (Becker, 1996) and (3) The Paul Scherrer Institut Smog Chamber (PSISC), CH (Paulsen et al., 2005). A brief technical description of each facility is given in Table 2.

2.3 Experiment Design

Table 2 provides a summary of the experiments conducted, which can be divided into three separate categories, (1) photooxidation, indoor chamber (Wyche et al., 2009;Alfarra et al., 2012;Alfarra et al., 2013), (2) photooxidation, outdoor chamber (Bloss et al., 2005;Camredon et al., 2010) and (3) mesocosm photooxidation, indoor chamber (Wyche et al., 2014). In each case the reaction chamber matrix comprised a temperature (T = 292 – 299 K) and humidity (49 – 84 % for photooxidation, indoor chamber and < 2 – 6 % for photooxidation, outdoor chamber) controlled synthetic air mixture. For all experiments the chamber air matrix also contained a pre-defined initial quantity of NO and NO_2 (VOC/NO_x ratios in the range 0.6 – 20, but typical ~ 2). The VOC precursor was introduced into the reaction chamber in liquid form *via* a heated inlet. In the case of the mesocosm photooxidation experiments, a known volume of air containing the precursor VOCs was transferred to the reaction chamber from a separate, illuminated plant chamber, which contained several tree specimens. For the indoor chamber systems, the experiments were initiated, after introduction of all reactants, by the switching on of artificial lights. For the outdoor chamber systems, the opening of the chamber cupola marked the start of the experiment. Experiments were typically run for 4 – 6 hours.

2.4 Instrumentation

CIR-TOF-MS was used to make *real-time* (*i.e.* 1 minute) measurements of the complex distribution of volatile organic compounds (ΣVOC, *i.e.* the sum of VOCs, oxygenated VOCs – OVOCs and nitrated VOCs – NVOCs) produced in the gas-phase during oxidation of each parent compound. In brief, the CIR-TOF-MS comprises a temperature controlled (T = 40 ^{o}C) ion source/drift cell assembly coupled to an orthogonal time-of-flight mass spectrometer equipped with a reflectron array (Kore Technology, UK). Proton Transfer Reaction (PTR) from hydronium (H_3O^+) and hydrated hydronium ($H_3O^+.(H_2O)_n$) was employed as the ionisation technique during all experiments (Jenkin et al., 2012). Further details regarding the CIR-TOF-MS can be found in Blake et al. (Blake et al., 2004) and Wyche et al. (Wyche et al., 2007).

Aerosol samples were collected on 47 mm quartz fibre filters at the end of certain experiments and the water-soluble organic content was extracted for analysis using LC-MS/MS. Reversed phase LC separation was achieved using an HP 1100 LC system equipped with an Eclipse ODS-C18 column with 5 µm particle size (Agilent, 4.6 mm × 150 mm). Mass spectrometric analysis was performed in negative ionisation mode using an HCT-Plus ion trap mass spectrometer with electrospray ionisation (Bruker Daltonics GmbH). Further details can be found in Hamilton et al. (Hamilton et al., 2003).

For several experiments, *real-time* broad chemical characterisation of the SOA was made using a cTOF-AMS (Aerodyne Research Inc., USA). The cTOF-AMS was operated in standard configuration, taking both mass spectrum (MS) and particle time-of-flight (PTOF) data; it was calibrated for ionisation efficiency using 350 nm monodisperse ammonium nitrate particles, the vapouriser was set to ~ 600 ^{o}C and a collection efficiency value of unity was applied (Alfarra et al., 2006). For further details, refer to Drewnick et al. (Drewnick et al., 2005) and Canagaratna et al. (Canagaratna et al., 2007).

Each chamber was additionally instrumented with on-line chemiluminescence/photolytic NO_2 NO_x analysers, UV photometric O_3 detectors, and scanning mobility particle sizers and condensation particle counters for aerosol size and number concentration, as well as temperature, pressure and humidity monitors. For full details regarding the various instrument suites employed at each chamber see Alfarra et al. (Alfarra et al., 2012), Paulsen et al. (Paulsen et al., 2005), Camredon et al. (Camredon et al., 2010) and references therein.

Filter and cTOF-AMS data were collected only during photooxidation experiments conducted at the MAC. Repeat experiments conducted at the MAC were carried out under similar starting conditions (*e.g.* VOC/NO_x ratio (Alfarra et al., 2013)).

2.5 Model construction

In order to aid analysis, the composition and evolution of the gas-phase components of the α-pinene chamber system were simulated using a chamber optimised photochemical box model incorporating the comprehensive α-pinene atmospheric oxidation scheme extracted from the Master Chemical Mechanism website (Jenkin et al., 1997;Saunders et al., 2003;Jenkin et al., 2012;http://mcm.leeds.ac.uk/MCM). The α-pinene mechanism employed (along with an appropriate inorganic reaction scheme) contained approximately 313 species and 942 different reactions. The box model employed also incorporated a series of "chamber specific" auxiliary reactions adapted from Bloss et al. (Bloss et al., 2005), Zador et al. (Zador et al., 2006) and Metzger et al., (Metzger et al., 2008) in order to take into account background chamber reactivity. Photolysis rates were parameterised for the PSI chamber and constrained using measured values of ($j(NO_2)$). All simulations were run at 295 K and 50 % relative humidity. NO, NO_2, HONO and α-pinene were either initialised or constrained, depending on the scenario investigated. For further details see Rickard *et al.* (Rickard et al., 2010).

3. Data Analysis

3.1 Data Processing

All CIR-TOF-MS data were recorded at a time resolution of 1-minute. In order to remove the time dimension and simultaneously increase detection limit, the individual mass spectra were integrated over the entire experiment; as such no account is taken of overall reaction time in the CIR-TOF-MS analysis. Removing the time dimension acts to reduce the dimensionality of the data, whilst maintaining the central characteristic spectral fingerprints produced by the photooxidation process. On average across all experiments studied, 98 % of the precursor had been consumed by the conclusion of the experiment; hence it is assumed that sufficient reaction took place in each instance to provide summed-normalised mass spectra that fully capture first- and higher-generation product formation.

The resultant summed spectra were normalised to 10^6 primary reagent ion counts (*i.e.* $\Sigma(H_3O^+ + H_3O^+.(H_2O)_n)$). Similarly normalised background spectra (recorded prior to injection of the precursor) were then subtracted from the summed-and-normalised experiment spectra. The 65 < *m/z* < 255 channels of the background removed spectra were extracted to comprise the region of interest. These ions tend to carry the most analyte-specific information, with lower *m/z* features tending to comprise either generic fragment ions that provide little chemical information (Blake et al., 2006) and/or small compounds emitted from illuminated chamber walls (*e.g.* (Bloss et al., 2005;Zador et al., 2006;Metzger et al., 2008). These extracted data were refined further by the application of a Mann-Whitney test (see Statistical Analysis for details), leaving residual spectra that comprised only the integrated-over-time signals corresponding to the VOC precursor and any reactive intermediate and product VOCs formed within the chamber during the experiment. Finally, the signal counts (in units of normalised counts per second; ncps) in each mass channel of the residuals, were expressed as a percentage of the total ion count in the refined region of interest.

The LC-MS/MS signal intensity data for the region 51 < *m/z* < 599 were extracted for analysis. For the AMS data, a 10-minute average was produced at 4 hours after lights on (roughly around the time when SOA mass had reached a peak and towards to the end of the experiment) and the region 40 < *m/z* < 150 (again the region carrying the most information; Alfarra et al., 2006) was extracted. Similar to the gas-phase data sets, the LC-MS/MS and AMS data were filtered using a Mann-Whitney test. Finally, for each data set all signal counts were expressed as a percentage of the total ion count in the respective *m/z* region of interest.

3.2 Statistical Analysis

Before any multivariate analysis was conducted, the processed CIR-TOF-MS, LC-MS/MS and AMS spectra were first filtered to remove unwanted data that were deemed to not be statistically significant. In order to do this, the mass spectra were initially grouped by structure of the precursor employed, giving seven separate groups for the CIR-TOF-MS data and three groups (owing to the smaller number of precursor species investigated) for the LC-MS/MS and AMS data, respectively. A two-sided Mann-Whitney test was then used to assess whether signals reported in individual mass channels were significantly different from the corresponding signals measured during a blank experiment. SPSS V20 (IBM, USA) was used for the analysis. A *p* value of *< 0.05* was considered statistically significant. The final summed-normalised and filtered spectra were then subjected to a series of multivariate statistical analysis techniques in order to probe the underlying chemical information. PLS-Toolbox (Eigenvector Research Inc., USA) operated in MatLab (Mathworks, USA; PLS-Tool Box) was used for the analysis.

To begin with, to reduce the data and identify similarities between the precursor oxidation systems, a PCA was conducted on the BVOC dataset and the model generated was then employed to map the reactivity of fig and birch tree mesocosm systems and to investigate the fit of a typical anthropogenic system (toluene) into the PCA space (both introduced into the model as test datasets). An *unsupervised pattern recognition*, hierarchical cluster analysis was also conducted on the data and a dendrogram produced to test relatedness, support the PCA and help interpret the precursor class separations achieved. The dendrogram was constructed using PCA scores, the centroid method and Mahalanobis distance coefficients. Finally, a *supervised pattern recognition* PLS-DA analysis was employed as a check for false-positives and as a quantitative classification tool to test the effectiveness of classification of the various systems in the model.

For the superposition of “classification” confidence limits onto the results of the PCA and HCA and for classification discrimination in the PLS-DA, prior to analysis the experiments were grouped according to the structure of the precursor investigated. Group 1 = isoprene (hemiterpene) and group 2 = α-pinene and limonene (both cyclic monoterpenes with an endocyclic double bond). Although limonene also has an exocyclic double bond in a side chain, we justify this classification on account of the endocyclic double bond in limonene being much more reactive towards ozone and slightly more reactive towards OH (Calvert et al., 2000). Group 3 = β-caryophyllene (sesquiterpene) and group 4 = myrcene (straight chain monoterpene) and linalool (straight chain OVOC). Strictly speaking, linalool is an OVOC (structure $C_{10}H_{18}O$) and not a monoterpene (structure $C_{10}H_{16}$), however we justify this grouping on account of both myrcene and linalool comprising primary BVOCs (often co-emitted; (Bouvier-

Brown et al., 2009;Kim et al., 2010;Wyche et al., 2014)) with certain structural similarities.

4. Results

4.1 Experiment overview

The temporal evolution of various key gas-phase (a) and particle-phase (b) parameters measured during a typical photooxidation experiment, are shown in Figure 1 in order to provide background context. In this instance the precursor was myrcene and the facility employed was the MAC. Full details describing the underlying chemical and physical mechanisms at play within such experiments can be found elsewhere (e.g. Larsen et al., 2001; Bloss et al., 2005; Paulsen et al., 2005; Surratt et al., 2006 and 2010; Wyche et al., 2009; 2014; Camredon et al., 2010; Rickard et al., 2010; Eddingsaas et al., 2012b; Hamilton et al., 2011; Jenkin et al., 2012; Alfarra et al., 2012, 2013; and references therein).

4.2 Mapping gas-phase composition

Of the 191 different mass channels extracted from the CIR-TOF-MS data for analysis (*i.e.* $65 < m/z < 255$), the Mann-Whitney test identified 151 as significant for one or more of the terpene precursor groups tested. These data were subsequently subjected to PCA. From inspection of the Eigenvalues derived, four principal components (PCs) were selected for analysis, which collectively accounted for 96 % of the variance within the data, with PCs 1 and 2 accounting for the vast majority, *i.e.* 63 and 18 %, respectively. This step, therefore, reduced the temporal traces of 191 mass-spectrum peaks to 4 composite and orthogonal dimensions.

Figure 2 shows a loadings bi-plot of PC2 *vs.* PC1. It is clear from Figure 2, that the model is able to successfully separate the four different classes of biogenic systems investigated. β-caryophyllene mass spectra are grouped in the upper left-hand quadrant of Figure 2, the monoterpenes in the lower left-hand quadrant and isoprene to the centre right. Moreover, the principal component analysis is able to distinguish between the cyclic monoterpene experiments of limonene and α-pinene (grouped into one class), and the straight chain monoterpene experiments of myrcene and linalool (grouped into a second class), albeit with the latter having a greater spread in confidence.

The *m/z* loadings of the PCA allow us to understand how the spectral fingerprints of the different terpene oxidation systems are grouped/separated by the PCA model. The first set of ions that contribute to separation of the different terpene systems comprises the protonated parent ions (MH^+) of the precursors themselves (and major fragments thereof), *i.e. m/z* 69 for isoprene, 137 (and fragment 81) for all monoterpenes (regardless of structure) and 205 for β-caryophyllene. Important contributions are to be expected from the respective parent-ions (being the basis for the use of chemical-ionisation mass spectrometry as an analyser of gas mixtures (Blake et al., 2009)). Our purpose here goes beyond identification of precursor and intermediate VOCs to an interpretation of reaction pathways in complex mixtures and potential linkages to SOA. In doing this, a certain amount of disambiguation of isobaric compounds becomes possible; indeed, as discussed in more detail below, Figure 2 clearly shows separation between cyclic and non-cyclic monoterpene oxidation groups, both of which have precursors of molecular weight (MW) 136 g mol^{-1}. Note, for clarity within Figure 2, the scale has been set to show the bulk of the data, hence precursor parent ions and *m/z* 71 are not shown.

Moving past the precursors into the detailed chemical information provided by the oxidation products formed within the chamber, we can see from Figure 2 that amongst others, *m/z* 71 (methyl vinyl ketone and methacrolein), 75 (hydroxy acetone), 83 (methyl furan) and 87 (C_4-hydroxycarbonyls/methacrylic acid) all contribute to separation of the isoprene group, and *m/z* 237 (β-caryophyllene aldehyde) and 235 and 253 (β-caryophyllene secondary ozonide and isomers thereof) to that of the β-caryophyllene group. The monoterpene groupings are influenced by the presence of *m/z* 107, 151 and 169 (primary aldehydes-piononaldehyde and limononaldehyde) and 139 (primary ketone- limonaketone) ions in their mass spectra. Helping to separate the straight chain from cyclic monoterpenes are *m/z* 95 and 93, relatively dominant features in both the myrcene and linalool spectra (relative abundance 10 – 24 % for *m/z* 93). *m/z* 93 has previously been identified as a major fragment ion of first generation myrcene and linalool products 4-vinyl-4-pentenal and 4-hydroxy-4-methyl-5-hexen-1-al, respectively (Shu et al., 1997;Lee et al., 2006). A list of major ions contributing to the separation of spectra in statistical space is given in Table 3 along with potential identities and precursors. It is worthy of note here that these ions and the overall fragmentation patterns observed in this study are largely in-line with those reported by Lee et al. (2006), in their comprehensive PTR-MS analysis of a wide range of BVOC precursors and their associated oxidation products.

4.3 Implementation of the model to classify mesocosm data

Having employed the terpene data as a training set to construct a PCA model, a test set of mesocosm data was introduced in order to investigate the ability of the model to map the classification of more complex biogenic mixtures. In this instance the mesocosm test set comprised two birch tree and two fig tree photooxidation experiments, containing a more complex and "realistic" mixture of various different VOCs (Wyche et al., 2014). The resultant scores plot is shown in Figure 3.

Figure 3 demonstrates that the model can successfully distinguish between the two different types of mesocosm systems. Moreover, the model correctly classifies the mesocosm systems within the PCA space, with the birch trees (which primarily emit monoterpenes and only small quantities of isoprene; (Wyche et al., 2014)) grouped with the single precursor monoterpene cluster, and the fig trees (which primarily emit isoprene and camphor and only a small amount of monoterpenes; (Wyche et al., 2014)) grouped between the monoterpene and isoprene clusters. Investigation of the mesocosm mass spectra and PCA loadings shows that mass channels 137, 139, 107, 95, 93, 81 and 71 are amongst features important in classifying the birch tree systems, with the relatively strong presence of *m/z* 93 suggesting the emission of noncyclic as well as cyclic monoterpenes from the birch trees. This was confirmed by cross-reference with GC-MS analysis, which showed that the acyclic monoterpene, ocimene, was the third most abundant monoterpene present in the birch tree emissions (Wyche et al., 2014). For the fig tree systems, mass channels 153, 81, 73, 71 and 69 are key for classification, with the presence of small quantities of camphor (*m/z* 153) and monoterpenes (*m/z* 81) causing the group to undergo a lateral shift in the PCA space, along PC1 away from the single precursor isoprene cluster.

As a further test of the technique to distinguish between and to classify VOCs and the matrix of oxidized organic compounds that may derive from their atmospheric chemistry, test data from an anthropogenic system was introduced into the model. In this instance, the toluene photooxidation system was employed. Toluene is an important pollutant in urban environments, originating from vehicle exhausts and fuel evaporation; furthermore it represents a model mono-aromatic, SOA precursor system (*e.g.* (Bloss et al., 2005)). As can be seen from the resultant scores plot in Figure 4, the model is also able to discriminate the anthropogenic system from those of biogenic origin. Besides the protonated toluene parent ion, those ions contributing to the positioning of the toluene cluster within the PCA space, include the protonated parent ions *m/z* 109 and 107, *i.e.* the ring retaining primary products benzaldehyde and phenol, respectively; *m/z* 123, *i.e.* the ring retaining secondary product, methyl benzoquinone and *m/z* 99 and 85, *i.e.* higher generation ring opening products (*e.g.* 4-oxo-2-pentenal and butenedial, respectively). A brief discussion regarding aromatic BVOCs is provided separately in the supplementary material.

4.4 Cluster analysis and classification

The relationships between the various terpene and mesocosm systems and their groupings with respect to one another can be explored further *via* the implementation of HCA; Figure 5 gives the dendrogram produced. Inspection of Figure 5 provides further evidence that the various systems in the four classes of terpenes investigated distinctly group together, with overall relatedness < 1 on the (centroid) distance between clusters scale using the Mahalanobis distance measure (Mahalanobis, 1936). Figure 5 shows that the sesquiterpene oxidation system has the most distinct spectral fingerprint (containing distinctive, higher mass oxidation products, *e.g. m/z* 253) and that the cyclic and straight chain monoterpene systems appear the most similar (with some common features alongside key, unique precursor/mechanism specific product patterns, *e.g. m/z* 93 for myrcene and linalool), grouping together with subclusters of cyclic and noncyclic precursors. The monoterpene dominated birch tree mesocosm experiments are grouped with the cyclic monoterpenes and show a close relationship with noncyclic monoterpene systems. Being dominated by isoprene emissions, yet with some monoterpenes and camphor present, the fig tree mesocosm experiments group separately but with a close degree of relation to the single precursor isoprene experiments.

In order to advance our chemometric mapping of biogenic systems beyond PCA and HCA (which do not consider user supplied *a priori* observation "class" information) and to provide a degree of quantification to our analysis, a PLS-DA using six latent variables (LVs) was conducted on the terpene and mesocosm data. For the PLS-DA, the experiments were grouped into their respective "classes", *i.e.* hemiterpene = isoprene; cyclic monoterpene = α-pinene and limonene; sesquiterpene = β-caryophyllene; noncyclic monoterpene = myrcene and linalool; birch trees; fig trees. Figure 6 shows a plot of the resultant scores on the first three LVs (accounting for ~ 85 % of the variance), from which it is clear that the PLS-DA is able to successfully discriminate between the four terpene classes, and places the monoterpene dominant birch experiments within the single precursor monoterpene cluster, and the isoprene dominant fig experiments close to the single precursor isoprene cluster within the PLS-DA model. The greater spread in confidence of the noncyclic monoterpene group is once again likely to result to some extend from the low number of repeat experiments employed (i.e. only two each for myrcene and linalool).

As can be seen from inspection of Table 4, model classification sensitivity and specificity was high in each instance. Each of the biogenic systems studied were predicted with 100 % sensitivity (with the exception of birch mesocosm), meaning that each set of experiments (again, except birch mesocosm) was predicted to fit perfectly within its class. The relatively low sensitivity obtained for birch mesocosm (50 %), is most likely a result of the use of only two repeat experiments in the model, coupled with experiment limitations and ageing trees producing slightly lower emissions during the final birch mesocosm experiment. All of the systems were predicted with > 90 % specificity (four of the six with 100 % specificity), indicating that all experiments are highly unlikely to be incorrectly classified.

4.5 Mapping particle-phase composition

In order to explore similar classifications and linkages in the concomitant particle-phase, the PCA, HCA and PLS-DA techniques were also applied to the off-line LC-MS/MS spectra obtained from analysis of filter samples and on-line AMS spectra.

As can be seen from inspection of Figure 7, the detailed LC-MS/MS aerosol spectra produce PCA results somewhat similar to those of the gas-phase CIR-TOF-MS spectra, with distinct clusters of cyclic monoterpenes, straight chain monoterpenes and sesquiterpenes. From inspection of the loadings components of the bi-plot (Figure 7a), we can see that *m/z* 237 (3-[2,2-dimethyl-4-(1-methylene-4-oxo-butyl)-cyclobutyl]-propanoic acid), 251 (β–caryophyllonic acid), 255 (4-(2-(2-carboxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid), 267 (β-14-hydroxycaryophyllonic acid and β-10-hydroxycaryophyllonic acid) and 271 (4-(2-(3-hydroperoxy-3-oxopropyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid or 4-(2-(2-carboxy-1-hydroxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid), are amongst those ions dominant in classifying the sesquiterpenes. For further details regarding β-caryophyllene oxidation products, see for example Lee et al. (2006), Winterhalter et al. (2009), Hamilton et al. (2011), Chan et al. (2011), Li et al. (2011) and Jenkin et al. (2012) and references therein, and Sect. 5. Of this set of oxidation products, β-caryophyllonic acid is common between the gas- (*i.e. m/z* 253) and particle- (*i.e. m/z* 251) phases.

Similarly, those ions (compounds) significant in isolating the cyclic monoterpenes include, *m/z* 169 (pinalic-3-acid, ketolimononaldehyde and limonalic acid), 183 (pinonic acid, limononic acid and 7-hydroxylimononaldehyde) and 185 (pinic acid, limonic acid), of which only those compounds of *m/z* 169 were observed to be of significant contribution to the gas-phase composition (observed as *m/z* 171; relative contribution as high as 1 - 5 % during α-pinene experiments). For further details regarding α-pinene and limonene oxidation products, see for example Larsen et al. (2001), Jaoui et al. (2003), Capouet et al. (2004), Jenkin (2004), Jaoui et al. (2006), Lee et al. (2006), Ng et al. (2006), Camredon et al. (2010) and Hamilton et al. (2011) and references therein. Comparatively little information is available on the speciated composition of myrcene and linalool SOA, however, from Figure 7a it is clear that somewhat larger mass compounds are important in classifying straight chain monoterpenes, *e.g. m/z* 321 (adduct ion $[M\text{-}H_2\text{+}FA\text{+}Na]^-$ M = 254 Da; potential formulae - $C_{12}H_{14}O_6$, six double bond equivalents or $C_{13}H_{18}O_5$, five double bond equivalents; indicative of oligomer formation), 325, 322 (the C13 peak for the *m/z* 321 ion), 227 ($C_{10}H_{11}O_6$), 215 ($C_{10}H_{15}O_5$) and 199 ($C_9H_{11}O_5$). Compounds of such high molecular weight were not observed in the concomitant gas-phase spectra. A list of major ions contributing to the separation of spectra in statistical space is given in Table 3 along with potential identities and precursors.

As with the PCA, the dendrogram produced *via* cluster analysis of the LC-MS/MS particle-phase data gave three distinct clusters (Figure 7b), *i.e.* cyclic monoterpene, straight chain monoterpene and sesquiterpene. The corresponding PLS-DA analysis reported 100 % sensitivity in each case and 100 % specificity for all systems except sesquiterpenes (*i.e.* β-caryophyllene = 83 %), suggesting a good level of model classification for the three types of terpene systems studied.

Despite utilising the somewhat destructive electron impact (EI) ionisation technique, the cTOF-AMS produces spectra of sufficient chemical detail such that the PCA and HCA are able to successfully differentiate between the groups of terpenes tested (Figure 8a and b). However, unlike the outputs from the CIR-TOF-MS and LC-MS/MS PCA's, the cyclic and straight chain monoterpenes in the AMS PCA do not group into two distinct classes, instead they tend to group in their species-specific sub-classes within the upper half of the PCA space. Indeed, the PLS-DA gave 100 % sensitivity and specificity for the cyclic monoterpenes and sesquiterpenes, but only 75 % sensitivity for the straight chain monoterpenes, suggesting that the model does less well at assigning myrcene and linalool cTOF-AMS spectra to their defined class.

As can be seen from inspection of Figure 8a, α-pinene, limonene and linalool tend in general to cluster towards the upper and right regions of the PCA space, primarily owing to the significant presence of *m/z* 43 and to a lesser extent *m/z* 44, in their spectra; both ions constituting common fragments observed in AMS of SOA (Alfarra et al., 2006). During such chamber experiments, the *m/z* 43 peak tends to comprise the CH_3CO^+ ion, originating from oxidised compounds containing carbonyl functionalities; it is usually representative of freshly oxidised material and semi-volatile oxygenated organic aerosol (SV-OOA; (Alfarra et al., 2006)).

From further inspection of the loadings bi-plot (Figure 8a) we see that the four sesquiterpene (β-caryophyllene) experiments cluster towards the lower left hand quadrant, their clustering heavily influenced by the presence of *m/z* 41 in their spectra as well as *m/z* 55, 79 and 95. In EI-AMS, *m/z* 41 comprises the unsaturated $C_3H_5^+$ fragment (Alfarra et al., 2006). As well as being influenced by the *m/z* 41 ion, the myrcene cluster (situated in the region of both the α-pinene and β-caryophyllene clusters in the PCA space) is also influenced by *m/z* 44, *i.e.* most likely the CO_2^+ ion. In this instance *m/z* 44 would tend to result from low volatility oxygenated organic aerosol (LV-OOA), derived from highly oxidised compounds, including oxo- and di-carboxylic acids (Alfarra et al., 2004;Alfarra et al., 2006). For full details regarding the particle-phase specific experiments conducted at the MAC, see Hamilton et al. (2011), Jenkin et al., (2012) and Alfarra et al. (2012 and 2013).

5. Discussion

5.1 Mapping chemistry

Figure 9 provides a highly simplified overview of the current state of knowledge regarding the atmospheric oxidation of hemi-, sesqui-, cyclic and straight chain mono-terpenes, showing selected key steps and intermediates on route to SOA formation. The mechanisms outlined in Figure 9 underpin the findings reported here and explain how the atmospheric chemistry of the various terpene oxidation systems and their SOA can be chemometrically mapped with respect to one another.

From a review of recent literature and from the summary presented in Figure 9, it can be seen that isoprene can react to form condensable second and higher generation nitrates in the presence of NO_x, *e.g.* C_4-hydroxy nitrate peroxy acetyl nitrate (C4-HN-PAN in Figure 9) (Surratt et al., 2010), as well as condensable OVOCs, *e.g.* hydroxymethyl-methyl-α-lactone (HMML) (Kjaergaard et al., 2012) and methacrylic acid epoxide (MAE) (Lin et al., 2013), *via* metharcolein (MACR) and methacryloyl-peroxy nitrate (MPAN). Alternatively, under "low NO_x" conditions (*e.g.* < 1 ppbV) isoprene can react to form condensable second-generation epoxides, *e.g.* isoprene epoxides (IEPOX), *via* primary peroxides (ISOPOOH) (Paulot et al., 2009a;Surratt et al., 2006)). Such C_4 and C_5 saturated, low volatility species constitute the monomer building blocks that proceed to form relatively high O:C ratio (nitrated in the presence of NO_x and sulphated in the presence of H_2SO_4) isoprene SOA oligomers (*e.g.* 2-methyl tetrol dimer O:C = 7:9) (Claeys et al., 2004;Surratt et al., 2006;Surratt et al., 2010;Worton et al., 2013). Consequently, the gas-phase composition under conditions forming isoprene SOA will therefore be dominated by relatively low MW monomer precursors, *e.g.* MACR (MH^+ = *m/z* 71), isoprene nitrates (ISOPN in Figure 9; MH^+ - HNO_3 = *m/z* 85) and MPAN ($MH^+.H_2O$ - HNO_3 = *m/z* 103) under "high NO_x" conditions (*e.g.* ~ 10's – 100's ppbV; (Paulot et al., 2009b;Surratt et al., 2010;Surratt et al., 2006)), and ISOPOOH and IEPOX (MH^+ - H_2O = *m/z* 101) under "low NO_x" conditions. For the "high NO_x" isoprene experiments conducted here, besides *m/z* 71, *i.e.* MACR (measured together with methyl vinyl ketone), *m/z* 87, 85, 83 and 75 *i.e.* (tentatively assigned to be) C_4-hydroxycarbonyls/methacrylic acid, ISOPN, C_5-hydroxy carbonyls (C5HC in Figure 9)/3-methyl furan (3-MF) and hydroxy acetone, respectively, were significant in classifying the isoprene group; MPAN at the *m/z* 103 ion was only a minor contributor. It should be noted that in theory, both HMML and MAE (MH^+ = *m/z* 103) may produce fragment ions of *m/z* 85 (*i.e.* MH^+-H_2O) following PTR ionisation, however without further detailed characterisation we are unable at this stage to postulate their fractional contribution to the measured *m/z* 85 signal.

Depending on the chemistry involved (Figure 9), potential SOA forming monoterpene products will either be (six-member-) ring retaining (*e.g.* from reaction with OH) or (six-member-) ring cleaved (*e.g.* from reaction with OH or O_3), producing gas-phase spectra with mid MW C_9 and C_{10} oxygenated (and nitrated in the presence of NO_x) products (*e.g.* (Kamens and Jaoui, 2001;Larsen et al., 2001;Capouet et al., 2004;Yu et al., 2008;Camredon et al., 2010;Eddingsaas et al., 2012b)). Both (six-member-) ring retaining and (six-member-) ring-opening products have been observed in monoterpene SOA (*e.g.* (Yu et al., 1999;Larsen et al., 2001;Camredon et al., 2010)), with the latter generally being dominant in terms of abundance (Camredon et al., 2010). Furthermore, (six-member-) ring-opening products are believed to undergo chemistry within the aerosol to form relatively low O:C ratio oligomers (*e.g.* 10-hydroxy-pinonic acid-pinonic acid dimmer, O:C = 7:19) (Gao et al., 2004;Tolocka et al., 2004;Camredon et al., 2010).

OH will react with straight chain monoterpenes, such as myrcene, primarily by addition to either the isolated or the conjugated double bond system. Reaction at the isolated C=C bond can proceed *via* fragmentation of the carbon backbone, producing acetone and mid MW, unsaturated C_7 OVOCs (and/or NVOCs, depending on NO_x levels). Reaction at the conjugated double bond system in myrcene would be expected to form formaldehyde in conjunction with either a C_9 aldehyde or C_9 ketone. Structure activity relationships (SARs) predict that the conjugated double bond system accounts for almost half of the OH reactivity. The conjugated double bond would therefore be expected to have a partial rate coefficient of the order 1×10^{-10} (*i.e.* similar to OH + isoprene) (Atkinson and Arey, 2003b). Consistent with this, the reported yields of acetone and formaldehyde from OH + myrcene are similar (Atkinson and Arey, 2003b), suggesting that the isolated double bond and the conjugated double bond system have comparable OH reactivity, as such we would expect C_9 and C_7 co-products to be formed in comparable yields. However, with a significant fraction of reactions with OH leading to the loss of three carbon atoms from the parent structure, the straight chain monoterpene gas-phase spectra tend to contain fewer features of MW greater than that of the precursor and more mid MW features. It tends to be these mid MW features, such as *m/z* 111 and 93 (*e.g.* 4-vinyl-4-pentenal, MYR 1.2 in Figure 9, MH^+ and MH^+-H_2O, respectively) and 113 and 95 (*e.g.* 2-methylenepentanedial MH^+ and MH^+-H_2O, respectively) that assist in the classification of the straight chain monoterpene experiments within the statistical space. Besides these ions, *m/z* 139 (primary myrcene C_9 aldehyde and/or C_9 ketone product) also assists in separating the myrcene spectra from those of α-pinene.

By comparing both the gas- and particle-phase cyclic monoterpenes in Figures 2 and 7a, it is evident that the dominant loadings represent compounds of similar MW, *i.e.* 169, 151 and 107 (primary aldehyde product, *e.g.* pinonaldehyde- PINAL in Figure 9, parent ion and fragments thereof) and 139 (primary ketone product parent ion) for the gas-phase and 187, 185, 183 and 169 for the particle-phase. Conversely, for the straight chain monoterpene experiments the major gas-phase loadings represent compounds of significantly smaller MW than their particle-phase counterparts, *i.e.* 113 and 95 and 111 and 93, compared to 325, 322, 321, 227 and 215. Indeed, the straight chain monoterpene LC-MS/MS spectra contained on average ~ 10 % more signal > 250 Da than the cyclic monoterpene spectra. Also, the composition of the ions observed in the straight chain monoterpene LC-MS/MS spectra suggests that the SOA particles contained both oligomers and highly oxidized species, with the C_{10} backbone intact (*i.e.* O:C = 0.6), similar in structure to (but a little less oxidised than) extremely low volatility organic vapours (ELV-VOC), which have been observed previously in significant yield from α-pinene and limonene (as well as 6-nonenal) ozonolysis chamber experiments in the absence of an OH scavenger, as well as boreal forests in Finland (Ehn et al., 2014). Further evidence to elucidate the type of SOA formed from the oxidation of straight chain monoterpenes can be obtained from investigation of the grouping of myrcene spectra in the cTOF-AMS PCA (Figure 8a). In the hour-4 cTOF-AMS PCA loadings bi-plot, we see that the grouping of the myrcene spectra is influenced somewhat by both *m/z* 41 and 44, indicating the presence of LV-OOA in the SOA, potentially a result of oligomerisation or further oxidative heterogeneous chemistry involving reaction at remaining C=C double bond sites.

β-caryophyllene readily forms particulate matter on oxidation (e.g. Jaoui et al., 2003; Lee et al., 2006; Winterhalter et al., 2009; Alfarra et al., 2012; Chen et al., 2012), with reaction predominantly at one of the two C=C sites (*e.g.* with OH or O_3, although O_3 attack occurs almost exclusively at the endocyclic double bond (Jenkin et al., 2012)), yielding relatively low vapour pressure, unsaturated and oxygenated primary products (Figure 9), which have significant affinity for the particle-phase (Jenkin et al., 2012). A further oxidation step involving the second C=C site can result in increased oxygen (and/or nitrogen, depending on NO_x conditions) content, yet with little, if any reduction in the original C number. As with the cyclic monoterpene PCAs, the CIR-TOF-MS and LC-MS/MS PCA bi-plots demonstrate similarities in terms of classifying β-caryophyllene oxidation and SOA formation with comparable MW species, *e.g.* primary products β-caryophyllon aldehyde (MW 236, BCAL in Figure 9) and β-caryophyllene secondary ozonide in the gas-phase (MW 252, BCSOZ in Figure 9), β-caryophyllonic acid (MW 252, C141CO2H in Figure 9) in both phases and secondary product β-nocaryophyllinic acid (MW 254, C131CO2H in Figure 9) in the particle-phase. In the hour-4 cTOF-AMS PCA scores plot, the myrcene and β-caryophyllene clusters are located adjacent to one another, with β-caryophyllene classification also influenced by the *m/z* 41 peak, which similar to myrcene SOA for example, is indicative of higher oxidized content (Alfarra et al., 2012), a result of either the partitioning of higher generation gas-phase products or heterogeneous oxidation of condensed first or second generation products.

5.2 Mapping within a class

Within the monoterpene group there is a small degree of separation between the limonene and α-pinene experiments, with three out of the four α-pinene experiments located to the upper and right region of the monoterpene cluster. This distribution/separation within the group may be a consequence of precursor-specific reaction pathways; for instance, although structurally similar, α-pinene and limonene react at somewhat different rates with respect to both OH and O_3 (Atkinson and Arey, 2003a). Over a fixed time period, such system reactivity will govern the degree of oxygenated content present within a closed analyte matrix and may facilitate the isolation of specific reaction pathways. Furthermore, the separation of such similar gas-phase precursors within a class cluster may help us to elucidate differences in resultant SOA yield and composition (*e.g.* limonene tends to have a larger SOA yield than α-pinene; Lee et al., 2006; Fry et al., 2014). It therefore may be possible with the use of larger and more detailed data sets, to employ loading information to determine the importance of certain products to SOA composition. However, additional data to those reported here would be required to fully test this hypothesis.

5.3 Mapping reactivity

In order to explore how the PCA technique can be used to investigate product distributions driven by certain starting conditions, a separate analysis was conducted on the five toluene experiments. In this instance we investigate the product distribution dependency on initial VOC/NO_x ratios. The VOC/NO_x ratios employed nominally represent "low", "medium" and "high" NO_x conditions, with values of roughly 11 (*i.e.* low NO_x, "NO_x-limited" ozone formation conditions – as determined from simulation chamber ozone isopleth plots, see Wagner et al., 2003; two experiments), 4 (*i.e.* moderate NO_x, two experiments) and 1 (*i.e.* high NO_x, "VOC-limited"; one experiment), respectively. The resultant PCA loadings bi-plot (produced using the methodology described in Section 3) is given in Figure 10.

From inspection of the PCA loadings bi-plot in Figure 10, it is clear that the toluene photooxidation spectra distribute in statistical space according to their respective initial VOC/NO_x ratios. Figure 10 shows the low NO_x, high VOC/NO_x ratio experiments grouped in the lower right-hand quadrant of the PCA space, principally influenced by loadings representing toluene (*m/z* 93 and 77, parent and fragment ions, respectively; note *m/z* 93 off-scale in Figure 10) and cresol (*m/z* 109). Summed spectra containing larger quantities of precursor would suggest the presence of a less reactive environment, which is the case here, where low NO_x levels in the NO_x limited regime, result in low [OH] (reduced radical cycling) and low [O_3] (less NO to NO_2 conversions) (see also Bloss et al., 2005). Similarly, the relatively large contribution from cresol to the low NO_x summed spectra, originates from a larger net cresol concentration across the experiment on account of low system reactivity (*i.e.* loss *via* reaction with OH).

The moderate NO_x, medium VOC/NO_x experiments group uniquely in the lower left-hand quadrant of the PCA space in Figure 10, principally on account of loadings representing benzaldehyde (*m/z* 107) and the ring-opening products, citraconic anhydride (*m/z* 113), 4-oxo-2-pentenal, maleic anhydride and/or angelicalactone (*m/z* 99) and methyl glyoxal (*m/z* 73). The greater abundance of higher generation, ring-opening products implies a more reactive environment (i.e. increased chemical processing) than that formed under low NO_x conditions. Larger net benzaldehyde concentrations originate from greater system reactivity and greater abundance of NO to fuel the RO_2 + NO reaction.

The high NO_x, low VOC/NO_x ratio experiment is sited in the left-hand half of Figure 10, on account of it possessing higher system reactivity (with respect to the low NO_x experiments) and the resultant greater proportion of ring-opening product ions (as the case for the moderate NO_x experiments). However, the low VOC/NO_x ratio experiment is uniquely displaced into the upper region of the PCA space owing to a large contribution from 2-butenedial and/or 2(5H)-furanone (*m/z* 85, off scale in Figure 10) to the summed spectra (the yields of both of which are likely to be important under high NO_x conditions, owing to reaction through the RO_2 + NO channel).

6. Atmospheric relevance and future directions

Having successfully used the mechanistic fingerprints in the chamber data to construct descriptive statistical models of the gas- and particle-phases, and having applied the methodology to map mesocosm environments, a next logical step would be to use this detailed chemical knowledge to investigate ambient VOC and SOA composition data in an attempt to help elucidate and deconvolve the important chemistry controlling the gas- and particle-phase composition of inherently more complex real world environments.

If ambient biogenic gas/particle composition spectra of unknown origin, uncertain speciated composition and/or a high level of detail and complexity were to be mapped onto the relevant statistical model (*i.e.* introduced as a separate test set), their resultant vector description in the statistical space would provide information regarding the type of precursors present and the underlying chemical mechanisms at play, as exemplified by the classifying of the mesocosm experiments by the fraction of isoprene, monoterpene and sesquiterpene chemistry in the experimental fingerprints. Furthermore, as shown by the mapping of toluene photooxidation experiments into a separate and distinct cluster, the methodology is potentially able to be robust with respect to other chemical compositions expected for a “real world” environment that is significantly impacted by both anthropogenic and biogenic emissions (*e.g.* Houston, USA and the Black Forest – Munich, DE). This capability is important when attempting to understand the complex interactions that exist between urban and rural atmospheres and when attempting to understand VOC and SOA source identification.

One potential problem in moving from simulation chamber data to “real world” systems, would be the applicability of using “static” experimental spectra (*i.e.* time averaged) to build a model to accept “dynamic” data, in which there would be potentially overlapping reaction coordinates and multiple precursor and radical sources.

In order to investigate the impact of a more dynamic system on the composition of the gas-phase matrix and hence on the composition of the spectra employed to build the model, a zero-dimensional chamber box model was constructed for the α-pinene system and operated under three different scenarios:

(1) *Basic chamber simulation*: α-pinene concentration constrained to measurements (initial concentration 124 ppbV); NO and NO_2 initialised according to measurements (31 and 41 ppbV, respectively).
(2) *Spiked chamber simulation*: α-pinene constrained as in (1), but profile duplicated to represent a fresh injection of the precursor (at the midpoint of the experiment) on top of the already evolving matrix; constant 10 ppbV HONO employed as NO and radical source.
(3) *Constant injection chamber simulation*: α-pinene and HONO constrained to constant values of 5 and 10 ppbV, respectively.

It should be noted here that the model runs are not idealised. The aim of these simulations is to provide systematically more complex chemical systems with which to compare and contrast a simulation representing the measured dataset. For work regarding the evaluation of the MCM with respect to single VOC precursor chamber experiments (including model-measurement intercomparison), see for example, Bloss et al., 2005 (toluene), Metzger et al., 2008 and Rickard et al., 2010 (1,3,5-TMB), Camredon et al., 2010 (α-pinene) and Jenkin et al., 2012 (β-caryophyllene).

The results of the three different model scenarios are given in Figure 11, mapped through to (*i.e.* integrated across the experiment) the resultant simulated mass spectra.

Figures 11a and b show the results from scenario (1). Figure 11a gives the evolution of the system over the molecular weight region of interest with time and Figure 11b gives the scenario summed "model mass spectra", *i.e.* the relative abundance of all simulated compounds within the gas-phase molecular weight region of interest (with relative contributions from isobaric species summed into a single "peak"). Scenario (1) and Figures 11a and b approximate the experimental data employed within this work and constitute the model base-case.

Figures 11c and d show the results from scenario (2). Figure 11c clearly shows the second α-pinene injection on top of the evolving matrix and the resultant system evolution. Figure 11d show the "difference model mass spectra" between scenarios (1) and (2), from which it can clearly be seen that there is very little difference between the spectra of the basic model and the "spiked" system. The difference in "mass channel" relative abundance (ΔMC) is generally ≤ 2 %, with the exceptions of MWs 168 and 186. MW 168 primarily comprises pinonaldehyde, with a ΔMC of around - 6 %; pinonaldehyde is a primary product and is slightly lower in relative abundance in scenario (2) owing to the longer reaction time employed and the greater proportion of pinonaldehyde reacted. MW 186 comprises a number of primary and secondary products and has a ΔMC of roughly + 3 %.

The results from model scenario (3) are given in Figures 11e and f. As with scenario (2), there is no dramatic difference between the simulated mass spectra of scenario (3) and the base-case scenario (1). In this instance ΔMC is generally ≤ ± 5 %, with the exceptions of MWs 136 and 168 and MWs 121 and 245. The relative abundance of the precursor is lower in this case on account of the constraining method employed and once again the relative abundance of pinonaldehyde is slightly lower due to the longer reaction time. MW 121 solely comprises PAN and MW 245 primarily comprises a C_{10} tertiary nitrate ($C_{10}H_{15}NO_6$, MCM designation: C106NO3). Both species are slightly elevated with respect to the base-case in scenario (3) owing to the longer reaction time and the continual input of OH and NO into the model in the form HONO.

Scenarios (2) and (3) represent complex mixtures with overlapping reaction coordinates, each one step closer to a "real world" case than scenario (1) and the chamber data employed within this work. However, despite the increase in complexity of the scenarios, both exhibit very little compositional difference to the base-case scenario and hence the chamber data employed in this work. These results give some confidence that despite being constructed from summed simulation chamber data, the statistical models employed here represents a solid framework onto which real atmosphere spectra could be mapped and interpreted.

A further step in increasing complexity and hence a further step towards the "real world" system, would be the addition of other (potentially unidentified) precursors to the simulation, which may be at different stages of oxidation or have passed through different reactive environments. Further increases in complexity, beyond the analysis discussed here, will form the focus of future work.

7. Conclusions

A chemometric dimension reduction methodology, comprising PCA, HCA and PLS-DA has been successfully applied for the first time to complex gas- and particle-phase composition spectra of a wide range of BVOC and mesocosm environmental simulation chamber photooxidation experiments. The results show that the oxidized gas-phase atmosphere (*i.e.* the integrated reaction coordinate) of each different structural type of BVOC can be classified into a distinct group according to the controlling chemistry and the products formed. Indeed, a potential major strength of the data analysis methodology described here, could lie in the decoding of mechanisms into pathways (i.e. separation within a group on account of different underlying chemistry) and consequently linking chemical pathways to precursor compounds. Furthermore, the methodology was similarly able to differentiate between the types of SOA particles formed by each different class of terpene, both in the detailed and broad chemical composition spectra. In concert, these results show the different SOA formation chemistry, starting in the gas-phase, proceeding to govern the differences between the various terpene particle compositions.

The ability of the methodology employed here to efficiently and effectively “data mine” large and complex datasets becomes particularly pertinent when considering that modern instrumentation/techniques produce large quantities of high-resolution temporal and speciated data over potentially long observation periods. Such statistical mapping of organic reactivity offers the ability to simplify complex chemical datasets and provide rapid and meaningful insight into detailed reaction systems comprising hundreds of reactive species. Moreover, the demonstrated methodology has the potential to assist in the evaluation of (chamber and real world) modelling results, providing easy to use, comprehensive observational metrics with which to test and evaluate model mechanisms and outputs and thus help advance our understanding of complex organic oxidation chemistry and SOA formation.

8. Acknowledgements

The authors gratefully acknowledge the UK Natural Environment Research Council (NERC) for funding the APPRAISE ACES consortium (NE/E011217/1) and the TRAPOZ project (NE/E016081/1); the EU-FP7 EUROCHAMP-2 programme for funding the TOXIC project (E2-2009-06-24-0001); the EU ACCENT Access to Infrastructures program for funding work at the PSI and the EU PEGASOS project (FP7-ENV-2010-265148) for funding used to support this work. Drs A. R. Rickard and M. R. Alfarra were supported by the NERC National Centre for Atmospheric Sciences (NCAS). The authors would like to thank the University of Leicester Atmospheric Chemistry group for assistance throughout all experiments, including Drs Alex Parker, Chris Whyte, Iain White and Timo Carr; co-workers at the University of Manchester for assistance with MAC experiments; co-workers from Fundacion CEAM, Drs Marie Camredon and Salim Alam for assistance with EUPHORE experiments and co-workers from the Laboratory of Atmospheric Chemistry smog chamber facility at the Paul Scherrer Institute (PSI) for assistance with PSISC experiments.

9. References

Alfarra, M. R., Coe, H., Allan, J. D., Bower, K. N., Boudries, H., Canagaratna, M. R., Jimenez, J. L., Jayne, J. T., Garforth, A. A., Li, S. M., and Worsnop, D. R.: Characterization of urban and rural organic particulate in the lower Fraser valley using two aerodyne aerosol mass spectrometers, Atmospheric Environment, 38, 5745-5758, 10.1016/j.atmosenv.2004.01.054, 2004.

Alfarra, M. R., Paulsen, D., Gysel, M., Garforth, A. A., Dommen, J., Prevot, A. S. H., Worsnop, D. R., Baltensperger, U., and Coe, H.: A mass spectrometric study of secondary organic aerosols formed from the photooxidation of anthropogenic and biogenic precursors in a reaction chamber, Atmospheric Chemistry and Physics, 6, 5279-5293, 2006.

Alfarra, M. R., Hamilton, J. F., Wyche, K. P., Good, N., Ward, M. W., Carr, T., Barley, M. H., Monks, P. S., Jenkin, M. E., Lewis, A. C., and McFiggans, G. B.: The effect of photochemical ageing and initial precursor concentration on the composition and hygroscopic properties of beta-caryophyllene secondary organic aerosol, Atmospheric Chemistry and Physics, 12, 6417-6436, 10.5194/acp-12-6417-2012, 2012.

Alfarra, M. R., Good, N., Wyche, K. P., Hamilton, J. E., Monks, P. S., Lewis, A. C., and McFiggans, G.: Water uptake is independent of the inferred composition of secondary aerosols derived from multiple biogenic VOCs, Atmospheric Chemistry and Physics, 13, 11769-11789, 10.5194/acp-13-11769-2013, 2013.

Alier, M., van Drooge, B. L., Dall'Osto, M., Querol, X., Grimalt, J. O., and Tauler, R.: Source apportionment of submicron organic aerosol at an urban background and a road site in Barcelona (Spain) during SAPUSS, Atmospheric Chemistry and Physics, 13, 10353–10371, doi:10.5194/acp-13-10353-2013, 2013.

Atkinson, R., and Arey: Gas phase tropospheric chemistry of biogenic volatile organic compounds- a review, Atmospheric Environment, 37, S197 - S219, 2003a.

Atkinson, R., and Arey, J.: Atmospheric Degradation of Volatile Organic Compounds, Chem. Rev., 103, 4605-4638, 2003b.

Baltensperger, U., Dommen, J., Alfarra, R., Duplissy, J., Gaeggeler, K., Metzger, A., Facchini, M. C., Decesari, S., Finessi, E., Reinnig, C., Schott, M., Warnke, J., Hoffmann, T., Klatzer, B., Puxbaum, H., Geiser, M., Savi, M., Lang, D., Kalberer, M., and Geiser, T.: Combined determination of the chemical composition and of health effects of secondary organic aerosols: The POLYSOA project, Journal of Aerosol Medicine and Pulmonary Drug Delivery, 21, 145-154, 10.1089/jamp.2007.0655, 2008.

Becker, K. H.: The European Photoreactor EUPHORE, Final Re-port to the European Commission, Bergische Universitat Wuppertal, Germany, Wuppertal, 1996.

Blake, R. S., Whyte, C., Hughes, C. O., Ellis, A. M., and Monks, P. S.: Demonstration of proton-transfer reaction time-of-flight mass spectrometry for real-time analysis of trace volatile organic compounds, Analytical Chemistry, 76, 3841-3845, 10.1021/ac0498260, 2004.

Blake, R. S., Wyche, K. P., Ellis, A. M., and Monks, P. S.: Chemical ionization reaction time-of-flight mass spectrometry: Multi-reagent analysis for determination of trace gas composition, International Journal Of Mass Spectrometry, 254, 85-93, 2006.

Blake, R. S., Monks, P. S., and Ellis, A. M.: Proton Transfer Reaction Mass Spectrometry, Chem. Rev., 109, 861-896, 2009.

Bloss, C., Wagner, V., Bonzanini, A., Jenkin, M. E., Wirtz, K., Martin-Reviejo, M., and Pilling, M. J.: Evaluation of detailed aromatic mechanisms (MCMv3 and MCMv3.1) against environmental chamber data, Atmospheric Chemistry and Physics 5, 623-639, 2005.

Bouvier-Brown, N. C., Goldstein, A. H., Gilman, J. B., Kuster, W. C., and de Gouw, J. A.: In-situ ambient quantification of monoterpenes, sesquiterpenes, and related oxygenated compounds during BEARPEX 2007: implications for gas- and particle-phase chemistry, Atmospheric Chemistry and Physics, 9, 5505-5518, doi:10.5194/acp-9-5505-2009, 2009.

Calogirou, A., Larsen, B. R., and Kotzias, D.: Gas-phase terpene oxidation products: a review, Atmospheric Environment, 33, 1423-1439, 10.1016/s1352-2310(98)00277-5, 1999.

Calvert, J. G., Atkinson, R., Kerr, J. A., Madronich, S., Moortgat, G. K., Wallington, T. J., and Yarwood, G.: The Mechanisms of Atmospheric Oxidation of the Alkenes, Oxford University Press, New York, 2000.

Camredon, M., Hamilton, J. F., Alam, M. S., Wyche, K. P., Carr, T., White, I. R., Monks, P. S., Rickard, A. R., and Bloss, W. J.: Distribution of gaseous and particulate organic composition during dark alpha-pinene ozonolysis, Atmospheric Chemistry and Physics, 10, 2893-2917, 2010.

Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q., Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R., Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols with the aerodyne aerosol mass spectrometer, Mass Spectrometry Reviews, 26, 185-222, :10.1002/mas.20115, 2007.

Capouet, M., Peeters, J., Noziere, B., and Muller, J. F.: Alpha-pinene oxidation by OH: simulations of laboratory experiments, Atmospheric Chemistry and Physics, 4, 2285 - 2311, 2004.

Chan, A. W. H., Kroll, J. H., Ng, N. L., and Seinfeld, J. H.: Kinetic modeling of secondary organic aerosol formation: effects of particle- and gas-phase reactions of semivolatile products, Atmos. Chem. Phys., 7, 4135–4147, doi:10.5194/acp-7-4135-2007, 2007.

Chan, M. N., Surratt, J. D., Chan, A. W. H., Schilling, K., Offenberg, J. H., Lewandowski, M., Edney, E. O., Kleindienst, T. E., Jaoui, M., Edgerton, E. S., Tanner, R. L., Shaw, S. L., Zheng, M., Knipping, E. M., and Seinfeld, J. H.: Influence of aerosol acidity on the chemical composition of secondary organic aerosol from β-caryophyllene. Atmos. Chem. Phys., 11, 1735–1751, doi:10.5194/acp-11-1735-2011, 2011.

Chen, Q., Li, Y. L., McKinney, K. A., Kuwata, M., and Martin, S. T.: Particle mass yield from β-caryophyllene ozonolysis, Atmos. Chem. Phys., 12, 3165-3179, doi: 10.5194/acp-12-3165-2012, 2012.

Chhabra, P. S., Ng, N. L., Canagaratna, M. R., Corrigan, A. L., Russell, L. M., Worsnop, D. R., Flagan, R. C., and Seinfeld, J. H.: Elemental composition and oxidation of chamber organic aerosol, Atmos. Chem. Phys., 11, 8827–8845, doi:10.5194/acp-11-8827-2011, 2011.

Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., Cafmeyer, J., Guyon, P., Andreae, M. O., Artaxo, P., and Maenhaut, W.: Formation of secondary organic aerosols through photooxidation of isoprene, Science, 303, 1173-1176, 2004.

Cocker III, D. R., Mader, B. T., Kalberer, M., Flagan, R. C., and Seinfeld, J. H.: The effect of water on gas-particle partitioning of secondary organic aerosol: II. m-xylene and 1,3,5-trimethylbenzene photooxidation systems, Atmos. Env., 35, 6073-6085, 2001.

Derwent, R. G., Jenkin, M. E., Passant, N. R., and Pilling, M. J.: Photochemical ozone creation potentials (POCPs) for different emission sources of organic compounds under European conditions estimated with Master Chemical Mechanism, Atmospheric Environment, 41, 2570-2579, 2007.

Dommen, J., Metzger, A., Duplissy, J., Kalberer, M., Alfarra, M. R., Gascho, A., Weingartner, E., Prevot, A. S. H., Verheggen, B., and Baltensperger, U.: Laboratory observation of oligomers in the aerosol from isoprene/NOx photooxidation, Geophys. Res. Lett., 33, L13805, doi:10.1029/2006gl026523, 2006.

Drewnick, F., Hings, S. S., DeCarlo, P., Jayne, J. T., Gonin, M., Fuhrer, K., Weimer, S., Jimenez, J. L., Demerjian, K. L., Borrmann, S., and Worsnop, D. R.: A new time-of-flight aerosol mass spectrometer (TOF-AMS) - Instrument description and first field deployment, Aerosol Sci. Technol., 39, 637-658, 10.1080/02786820500182040, 2005.

Eddingsaas, N. C., Loza, C. L., Yee, L. D., Seinfeld, J. H., and Wennberg, P. O.: alpha-pinene photooxidation under controlled chemical conditions - Part 1: Gas-phase composition in low- and high-NOx environments, Atmospheric Chemistry and Physics, 12, 6489-6504, 10.5194/acp-12-6489-2012, 2012a.

Eddingsaas, N. C., Loza, C. L., Yee, L. D., Chan, M., Schilling, K. A., Chhabra, P. S., Seinfeld, J. H., and Wennberg, P. O.: alpha-pinene photooxidation under controlled chemical conditions - Part 2: SOA yield and composition in low- and high-NOx environments, Atmospheric Chemistry and Physics, 12, 7413-7427, 10.5194/acp-12-7413-2012, 2012b.

Ehn, M., Thornton, J. A., Kleist, E., Sipila, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I. H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurten, T., Nielsen, L. B., Jorgensen, S., Kjaergaard, H. G., Canagaratna, M., Dal Maso, M., Berndt, T., Petaja, T., Wahner, A., Kerminen, V. M., Kulmala, M., Worsnop, D. R., Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol, Nature, 506, 476-+, 10.1038/nature13032, 2014.

Fantechi, G., Vereecken, L., and Peeters, J.: The OH initiated atmospheric oxidation of pinonaldehyde: Detailed theoretical study and mechanism construction, Phyical Chemistry Chemical Physics, 4, 5795 - 5805, 2002.

Frey, J. L., Draper, D. C., Barsanti, K. C., Smith, J. N., Ortega, J., P. M., Winkler, Lawler, M. J., Brown, S. S., Edwards, P. M., Cohen, R. C., and Lee, L.: Secondary Organic Aerosol Formation and Organic Nitrate Yield from NO_3 Oxidation of Biogenic Hydrocarbons, Environmental Science & Technology, 48(20), 11944-11953, 10.1021/es502204x, 2014

Gao, S., Ng, N. L., Keywood, M., Varutbangkul, V., Bahreini, R., Nenes, A., He, J., Yoo, K. Y., Beauchamp, J. L., Hodyss, R. P., Flagan, R. C., and Seinfeld, J. H.: Particle Phase Acidity and Oligomer Formation in Secondary Organic Aerosol, Environmental Science & Technology, 38, 6582-6589, 2004.

Glasius, M., Lahaniati, M., Calogirou, A., Bella, D. D., Jensen, N. R., Hjorth, J., Kotzias, D., and Larsen, B. R.: Carboxylic acids in secondary aerosol from oxidation of cyclic monoterpenes by ozone, Environ. Sci. Technol., 34, 1001–1010, 2000.

Grieshop, A. P., Donahue, N. M., and Robinson, A. L.: Is the gas-particle partitioning in alpha-pinene secondary organic aerosol reversible?, Geophys. Res. Lett., 34, L14810, doi:10.1029/2007GL029987, 2007.

Griffin, R. J., Cocker, D. R., III, Flagan, R. C., and Seinfeld, J. H.: Organic aerosol formation from the oxidation of biogenic hydrocarbons, J. Geophys. Res., 104, 3555–3567, 1999.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prevot, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmospheric Chemistry and Physics, 9, 5155-5236, 2009.

Hamilton, J. F., Lewis, A. C., Bloss, C., Wagner, V., Henderson, A. P., Golding, B. T., Wirtz, K., Martin-Reviejo, M., and Pilling, M. J.: Measurments of photo-oxidation products from the reaction of a series of alkyl-benzenes with hydroxyl radicals during EXACT using comprehensive gas chromatography, Atmospheric Chemistry and Physics, 3, 1999-2014, 2003.

Hamilton, J. F., Alfarra, M. R., Wyche, K. P., Ward, M. W., Lewis, A. C., McFiggans, G. B., Good, N., Monks, P. S., Carr, T., White, I. R., and Purvis, R. M.: Investigating the use of secondary organic aerosol as seed particles in simulation chamber experiments, Atmospheric Chemistry and Physics, 11, 5917-5929, 10.5194/acp-11-5917-2011, 2011.

Harrison, R. M., Giorio, C., Beddows, D. C. S., and Dall'Osto, M.: Size distribution of airborne particles controls outcome of epidemiological studies, Science of the Total Environment, 289-293, 2010.

Heal, M. R., Kumar, P., and Harrison, R. M.: Particles, air quality, policy and health, Chemical Society Reviews, 6606-6630, 10.1039/c2cs35076a, 2012.

Hennigan, C. J., Miracolo, M. A., Engelhart, G. J., May, A. A., Presto, A. A., Lee, T., Sullivan, A. P., McMeeking, G. R., Coe, H., Wold, C. E., Hao, W. –M., Gilman, J. B., Kuster, W. C., de Gouw, J., Schichtel, B. A., Collet Jr, J. L., Kreidenweis, S. M., and Robinson, A. L.: Chemical and physical transformations of organic aerosol from the photo-oxidation of open biomass burning emissions in an environmental chamber, Atmos. Chem. Phys., 11, 7669-7686, 2011.

Henze, D. K., and Seinfeld, J. H.: Global Secondary Organic Aerosol from isoprene oxidation, Geophysical Research Letters, 33, L09812, 2006.

Heringa, M. F., DeCarlo, P. F., Chirico, R., Tritscher, T., Clairotte, M., Mohr, C., Crippa, M., Slowik, J. G., Pfaffenberger, L., Dommen, J., Weingartner, E., Prevot, A. S. H., and Baltensperger, U.: A new method to discriminate secondary organic aerosols from different sources using high-resolution aerosol mass spectra, Atmospheric Chemistry and Physics, 12, 2189–2203, doi:10.5194/acp-12-2189-2012, 2012.

Hildebrandt, L., Donahue, N. M., and Pandis, S. N.: High formation of secondary organic aerosol from the photo-oxidation of toluene, Atmos. Chem. Phys., 9, 2973–2986, doi:10.5194/acp-9-2973-2009, 2009.

Hoffmann, T., Odum, J. R., Bowman, F., Collins, D., Klockow, D., Flagan, R. C., and Seinfeld, J. H.: Formation of Organic Aerosols from the Oxidation of Biogenic Hydrocarbons, J. Atmos. Chem., 26, 189-222, 1997.

http://mcm.leeds.ac.uk/MCM: in.

IPCC: Intergovernmental Panel on Climate Change, Fourth Assessment Report, Climate Change 2007, 2007.

Jackson, J. E.: PRINCIPAL COMPONENTS AND FACTOR-ANALYSIS .1. PRINCIPAL COMPONENTS, J. Qual. Technol., 12, 201-213, 1980.

Jaoui, M., and Kamens, R. M.: Mass balance of gaseous and particulate products analysis from alpha-pinene/NOx/air in the presence of natural sunlight, Journal of Geophysical Research-Atmospheres, 106, 12541-12558, 10.1029/2001jd900005, 2001.

Jaoui, M. and Kamens, R. M.: Gaseous and particulate oxidation products analysis of a mixture of a α-pinene + β-pinene/O3/air in the absence of light and α-pinene + β-pinene/NOx/air in the presence of natural sunlight, J. Atmos. Chem., 44, 259–297, 2003.

Jaoui, M., Leungsakul, S., and Kamens, R. M.: Gas and particle products distribution from the reaction of beta-caryophyllene with ozone, J. Atmos. Chem., 45, 261–287, 2003.

Jaoui, M., Corse, E., Kleindienst, T. E., Offenburg, J. H., Lewandowski, M., and Edney, E. O.: Analysis of Secondary Organic Aerosol Compounds from the Photooxidation of *d*-Limonene in the Presence of NO_x and their Detection in Ambient $PM_{2.5}$, Environ. Sci. Technol., 40(12), 3819-3828, doi: 10.1021/es052566z, 2006.

Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: a protocol for mechanism developement, Atmospheric Environment, 3, 81 - 104, 1997.

Jenkin, M. E.: Modelling the formation and composition of secondary organic aerosol from alpha- and beta-pinene ozonolysis using MCM v3, Atmospheric Chemistry and Physics, 4, 1741-1757, 2004.

Jenkin, M. E., Wyche, K. P., Evans, C. J., Carr, T., Monks, P. S., Alfarra, M. R., Barley, M. H., McFiggans, G. B., Young, J. C., and Rickard, A. R.: Development and chamber evaluation of the MCM v3.2 degradation scheme for beta-caryophyllene, Atmospheric Chemistry and Physics, 12, 5275-5308, 10.5194/acp-12-5275-2012, 2012.

Kalberer, M., Sax, M., and Samburova, V.: Molecular Size Evolution of Oligomers in Organic Aerosols Collectes in Urban Atmospheres and Generated in a Smog Chamber, Environmental Science & Technology, 40, 5917 - 5922, 2006.

Kamens, R. M., and Jaoui, M.: Modelling aerosol formation from a-pinene + NOx in the presence of natural sunlight using gas-phase kinetics and gas-particle partitioning theory, Environmental Science & Technology, 35, 1394 - 1405, 2001.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Detener, F. J., Facchini, M. C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. L., Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global climate modelling: a review, Atmospheric Chemistry and Physics, 5, 1053-1123, 2005.

Khamaganov, V. G., and Hites, R. A.: Rate Constants for the Gas-Phase Reactions of Ozone with Isoprene, α- and β-Pinene, and Limonene as a Function of Temperature, The Journal of Physical Chemistry A, 105, 815-822, DOI: 10.1021/jp002730z, 2001.

Kim, S., Karl, T., Guenther, A., Tyndall, G., Orlando, J., Harley, P., Rasmussen, R., and Apel, E.: Emissions and ambient distributions of Biogenic Volatile Organic Compounds (BVOC) in a ponderosa pine ecosystem: interpretation of PTR-MS mass spectra, Atmospheric Chemistry and Physics, 10, 1759-1771, doi:10.5194/acp-10-1759-2010, 2010.

Kjaergaard, H. G., Knap, H. C., Ornso, K. B., Jorgensen, S., Crounse, J. D., Paulot, F., and Wennberg, P. O.: Atmospheric Fate of Methacrolein. 2. Formation of Lactone and Implications for Organic Aerosol Production, Journal of Physical Chemistry A, 116, 5763-5768, 10.1021/jp210853h, 2012.

Kleindienst, T. E., Conver, T. S., McIver, C. D., and Edney, E. O.: Determination of secondary organic aerosol products from the photooxidation of toluene and their implications in ambient PM2.5, J. Atmos. Chem., 47, 79–100, 2004.

Koch, S., Winterhalter, R., Uherek, E., Kolloff, A., Neeb, P., and Moortgat, G. K.: Formation of new particles in the gas phase ozonolysis of monoterpenes, Atmospheric Environment, 34, 4031 - 4042, 2000.

Kroll, J. H., Ng, N. L., Murphy, S. M., Flagan, R. C., and Seinfeld, J. H.: Secondary organic aerosol formation from isoprene photooxidation under high-NO_x conditions, Geophysical Research Letters, 32, L18808, 2005.

Kroll, J. H., Ng, N. L., Murphy, S. M., Flagan, R. C., and Seinfeld, J. H.: Secondary organic aerosol formation from isoprene photooxidation, Environmental Science and Technology, 40, 1869 - 1877, 2006.

Kroll, J. H., and Seinfeld, J. H.: Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, Atmospheric Environment, 42, 3593-3624, 2008.

Kuppusami, S., Clokie, M. R. J., Panayi, T., Ellis, A. M., and Monks, P. S.: Metabolite profiling of Clostridium difficile ribotypes using small molecular weight volatile organic compounds, Metabolomics, DOI 10.1007/s11306-014-0692-4, 2014.

Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot, A. S. H.: Source apportionment of submicron organic aerosols at an urban site by factor analytical modelling of aerosol mass spectra, Atmos Chem Phys, 7, 1503-1522, 2007.

Larsen, B. R., Di Bella, D., Glasius, M., Winterhalter, R., Jensen, N. R., and Hjorth, J.: Gas-phase OH oxidation of monoterpenes: gaseous and particulate products, Journal of Atmospheric Chemistry, 38, 231 - 276, 2001.

Lee, A., Goldstein, A. H., Kroll, J. H., Ng, N. L., Varutbangkul, V., Flagan, R. C., and Seinfeld, J. H.: Gas-phase products and secondary aerosol yields from the photooxidation of 16 different terpenes, Journal of Geophysical Research-Atmospheres, 111, 10.1029/2006jd007050, 2006.

Li, Y. J., Chen, Q., Guzman, M. I., Chan, C. K., and Martin, S. T.: Second-generation products contribute substantially to the particle-phase organic material produced by β-caryophyllene ozonolysis, Atmos. Chem. Phys., 11, 121–132, doi:10.5194/acp-11-121-2011, 2011.

Librando, V., and Tringali, G.: Atmospheric fate of OH initiated oxidation of terpenes. Reaction mechanism of a-pinene degradation and secondary organic aerosol formation, Journal of Environmental Management, 75, 275 - 282, 2005.

Limbeck, A., Kulmala, M., and Puxbaum, H.: Secondary organic aerosol formation in the atmosphere via heterogeneous reaction of gaseous isoprene on acid particles, Geophysical Research Letters, 30, 2003.

Lin, Y.-H., Zhanga, H., Pye, H. O., Zhanga, Z., Martha, W. J., Parka, S., Arashiroa, M., Cuia, T., Budisulistiorinia, S. H., Sextona, K. G., Vizuetea, W., Xieb, Y., Lueckenb, D. J., Pileticb, I. R., Edneyb, E. O., Bartolottic, L. J., Gold, A., and Surratt, S. D.: Epoxide as a precursor to secondary organic aerosol formation from isoprene photooxidation in the presence of nitrogen oxides, Environmental Science & Technology, 110, 6718-6723, doi/10.1073/pnas.1221150110, 2013.

Mahalanobis, P. C.: On the generalised distance in statistics, Proceedings of the National Institute of Sciences of India, 2, 49-55, 1936.

Marcolli, C., Canagaratna, M. R., Worsnop, D. R., Bahreini, R., de Gouw, J. A.,Warneke, C., Goldan, P. D., Kuster,W. C.,Williams, E. J., Lerner, B. M., Roberts, J. M., Meagher, J. F., Fehsenfeld, F. C., Marchewka, M., Bertman, S. B., and

Middlebrook, A. M.: Cluster Analysis of the Organic Peaks in Bulk Mass Spectra Obtained During the 2002 New England Air Quality Study with an Aerodyne Aerosol Mass Spectrometer, Atmos. Chem. Phys., 6, 5649–5666, 2006, http://www.atmos-chem-phys.net/6/5649/2006/.

Metzger, A., Dommen, J., Gaeggeler, K., Duplissy, J., Prevot, A. S. H., Kleffmann, J., Elshorbany, Y., Wisthaler, A., and U., B.: Evaluation of 1,3,5-trimethylbenzene degredation in the detailed tropospheric chemistry mechanism MCMv3.1 using environmental chamber data, Atmospheric Chemistry and Physics, 8, 6453-6468, 2008.

Ng, N. L., Kroll, J. H., Keywood, M. D., Bahreini, R., Varutbangkul, V., Flagan, R. C., Seinfeld, J. H., Lee, A., and Goldstein, A. H.: Contribution of first- versus second-generation products to secondary organic aerosols formed in the oxidation of biogenic hydrocarbons, Environ. Sci. Technol., 40, 2283–2297, 2006.

Ng, N. L., Kroll, J. H., Chan, A. W. H., Chhabra, P. S., Flagan, R. C., and Seinfeld: Secondary organic aerosol formation from *m*-xylene, toluene, and benzene, Atmos. Chem. Phys., 7, 3909–3922, 2007.

Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., and Seinfeld, J. H.: Gas/particle partitioning and secondary organic aerosol yields, Environ. Sci. Technol., 30, 2580–2585, 1996.

Paatero, P. and Tapper, U.: Positive Matrix Factorization: a nonnegative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 111–126, 1994.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometr. Intell. Lab., 37, 23–35, 1997.

Paglione, M., Kiendler-Scharr, A., Mensah, A. A., Finessi, E., Giulianelli, L., Sandrini, S., Facchini, M. C., Fuzzi, S., Schlag, P., Piazzalunga, A., Tagliavini, E., Henzing, J. S., and S., D.: Identification of humic-like substances (HULIS) in oxygenated organic aerosols using NMR and AMS factor analyses and liquid chromatographic techniques, Atmospheric Chemistry and Physics, 14, 25-45, doi:10.5194/acp-14-25-2014, 2014.

Pandis, S. N., Paulson, S. E., Seinfeld, J. H., and Flagan, R. C.: Aerosol Formation in the Photooxidation of Isoprene and Beta-Pinene, Atmos. Environ., 25(5-6), 997–1008, doi: 10.1016/0960-1686(91)90141-S, 1992.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kurten, A., St Clair, J. M., Seinfeld, J. H., and Wennberg, P. O.: Unexpected Epoxide Formation in the Gas-Phase Photooxidation of Isoprene, Science, 325, 730-733, 10.1126/science.1172910, 2009a.

Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kroll, J. H., Seinfeld, J. H., and Wennberg, P. O.: Isoprene photooxidation: new insights into the production of acids and organic nitrates, Atmospheric Chemistry and Physics, 9, 1479-1501, 2009b.

Paulot, F., Henze, D. K., and Wennberg, P. O.: Impact of the isoprene photochemical cascade on tropical ozone, Atmospheric Chemistry and Physics, 12, 1307-1325, 10.5194/acp-12-1307-2012, 2012.

Paulsen, D., Dommen, J., Kalberer, M., Prevot, A. S. H., Richter, R., Sax, M., Steinbacher, M., Weingartner, E., and Baltensperger, U.: Secondary organic aerosol formation by irradiation of 1,3,5-trimethylbenzene-NOx-H2O in a new reaction chamber for atmospheric chemistry and physics, Environmental Science & Technology, 39, 2668-2678, 2005.

Presto, A. A., HuffHartz, K. E., and Donahue, N. M.: Secondary Organic Aerosol Production from Terpene Ozonolysis. 2. Effect of NOx Concentration, Environ. Sci. Technol., 39, 7046–7054, 2005.

Reinnig, M. C., Muller, L., Warnke, J., and Hoffmann, T.: Characterization of selected organic compound classes in secondary organic aerosol from biogenic VOCs by HPLC/MS, Analytical Bioanalytical Chemistry, 391, 171 - 182, 2008.

Rohrer, F., Bohn, B., Brauers, T., Bruning, 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, 2005.

Rickard, A. R., Wyche, K. P., Metzger, A., Monks, P. S., Ellis, A. M., Dommen, J., Baltensperger, U., Jenkin, M. E., and Pilling, M. J.: Gas phase precursors to anthropogenic secondary organic aerosol Using the Master Chemical Mechanism to probe detailed observations of 1,3,5-trimethylbenzene photo-oxidation, Atmospheric Environment, 44, 5423-5433, 10.1016/j.atmosenv.2009.09.043, 2010.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the developement of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degredation of non-aromatic volatile organic compounds, Atmospheric Chemistry and Physics, 3, 161 - 180, 2003.

Shu, Y. H., Kwok, E. S. C., Tuazon, E. C., Atkinson, R., and Arey, J.: Products of the gas-phase reactions of linalool with OH radicals, NO3 radicals, and O-3, Environmental Science & Technology, 31, 896-904, 10.1021/es960651o, 1997.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Muller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years, Atmospheric Chemistry and Physics, 14, 9317-9341, 10.5194/acp-14-9317-2014, 2014.

Slowik, J. G., Vlasenko, A., McGuire, M., Evans, G. J., and Abbatt, J. P. D.: Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site, Atmos. Chem Phys., 10, 1969-1988, 2010

Sousa, C., Novais, A., Magalhaes, A., Lopes, J., and Peixe, L.: Diverse high-risk B2 and D Escherichia coli clones depicted by Fourier Transform Infrared Spectroscopy, Sci Rep, 3, 8, 327810.1038/srep03278, 2013.

Sun, X., Zhang, C., Zhao, Y., Bai, J., and Hea, M.: Kinetic study on the linalool ozonolysis reaction in the atmosphere, Canadian Journal of Chemistry, 90, doi: 10.1139/v2012-001, 2012.

Surratt, J. D., Murphy, S. M., Kroll, J. H., Ng, N. L., Hilderbrandt, L., Sorooshian, A., Szmigielski, R., Vermeylen, R., Maenhaut, W., Claeys, M., Flagen, R., and Seinfeld, J. H.: Chemical Composition of Secondary Organic Aerosol Formed from the Photooxidation of Isoprene, Journal of Physical Chemistry A, 110, 9665-9690, 2006.

Surratt, J. D., Chan, A. W. H., Eddingsaas, N. C., Chan, M. N., Loza, C. L., Kwan, A. J., Hersey, S. P., Flagan, R. C., Wennberg, P. O., and Seinfeld, J. H.: Reactive intermediates revealed in secondary organic aerosol formation from isoprene, Proc. Natl. Acad. Sci. U. S. A., 107, 6640-6645, 10.1073/pnas.0911114107, 2010.

Surratt, J. D.: Radical regeneration from isoprene, Nat. Geosci., 6, 995-996, 2013.

Tolocka, M. P., Jang, M., Ginter, J. M., Cox, F. J., Kamens, R. M., and Johnston, M. V.: Formation of Oligomers in Secondary Organic Aerosol, Environmental Science & Technology, 38, 1428-1434, 10.1021/es035030r, 2004.

Wagner, V., Jenkin, M. E., Saunders, S. M., Stanton, J., Wirtz, K., and Pilling, M. J.: Modelling of the photooxidation of toluene: conceptual validating detailed mechanisms, Atmos. Chem. Phys., 3, 89-106, 2003.

Wang, J. L., Chew, C., Chang, C. Y., Liao, W. C., Lung, S. C. C., Chen, W. N., Lee, P. J., Lin, P. H., and Chang, C. C.: Biogenic isoprene in subtropical urban settings and implications for air quality, Atmospheric Environment, 79, 369-379, 10.1016/j.atmosenv.2013.06.055, 2013.

Winterhalter, R., Herrmann, F., Kanawati, B., Nguyen, T. L., Peeters, J., Vereecken, L., and Moortgat, G. K.: The gas-phase ozonolysis of β-caryophyllene (C15H24). Part I: an experimental study, Phys. Chem. Chem. Phys., 11, 4152–4172, 2009.

Worton, D. R., Surratt, J. D., LaFranchi, B. W., Chan, A. W. H., Zhao, Y. L., Weber, R. J., Park, J. H., Gilman, J. B., de Gouw, J., Park, C., Schade, G., Beaver, M., St Clair, J. M., Crounse, J., Wennberg, P., Wolfe, G. M., Harrold, S., Thornton, J. A., Farmer, D. K., Docherty, K. S., Cubison, M. J., Jimenez, J. L., Frossard, A. A., Russell, L. M., Kristensen, K., Glasius, M., Mao, J. Q., Ren, X. R., Brune, W., Browne, E. C., Pusede, S. E., Cohen, R. C., Seinfeld, J. H., and Goldsteint, A. H.: Observational Insights into

Aerosol Formation from Isoprene, Environmental Science & Technology, 47, 11403-11413, 10.1021/es4011064, 2013.

Wyche, K. P., Blake, R. S., Ellis, A. M., Monks, P. S., Brauers, T., Koppmann, R., and Apel, E. C.: Technical note: Performance of Chemical Ionization Reaction Time-of-Flight Mass Spectrometry (CIR-TOF-MS) for the measurement of atmospherically significant oxygenated volatile organic compounds, Atmospheric Chemistry and Physics, 7, 609-620, 2007.

Wyche, K. P., Monks, P. S., Ellis, A. M., Cordell, R. L., Parker, A. E., Whyte, C., Metzger, A., Dommen, J., Duplissy, J., Prevot, A. S. H., Baltensperger, U., Rickard, A. R., and Wulfert, F.: Gas phase precursors to anthropogenic secondary organic aerosol: detailed observations of 1,3,5-trimethylbenzene photooxidation, Atmospheric Chemistry and Physics, 9, 635-665, 2009.

Wyche, K. P., Ryan, A. C., Hewitt, C. N., Alfarra, M. R., McFiggans, G. B., Carr, T., Monks, P. S., Smallbone, K. L., Capes, G., Hamilton, J. F., Pugh, T. A. M., and MacKenzie, A. R.: Emissions of biogenic volatile organic compounds and subsequent photochemical production of secondary organic aerosol in mesocosm studies of temperate and tropical plant species, Atmospheric Chemistry and Physics, In Press, 2014.

Yu, J. Z., Cocker, D. R., Griffin, R. J., Flagan, R. C., and Seinfeld, J. H.: Gas-phase ozone oxidation of monoterpenes: Gaseous and particulate products, Journal of Atmospheric Chemistry, 34, 207-258, 10.1023/a:1006254930583, 1999.

Yu, Y., Ezell, M. J., Zelenyuk, A., Imre, D., Alexander, L., Ortega, J., D'Anna, B., Harmon, C. W., Johnson, S. N., and Finlayson-Pitts, B. J.: Photooxidation of alpha-pinene at high relative humidity in the presence of increasing concentrations of NOx, Atmospheric Environment, 42, 5044-5060, 10.1016/j.atmosenv.2008.02.026, 2008.

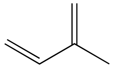
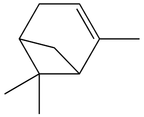
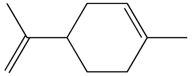
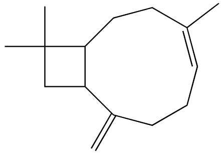
Zador, J., Turanyi, T., Wirtz, K., and Pilling, M. J.: Measurement and investigation of chamber radical sources in the EUropean PHoto REactor (EUPHORE), Journal of Atmospheric Chemistry, 55, 147 - 166, 2006.

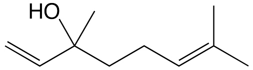
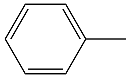
Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and Jimenez, J. L.: Deconvolution and Quantification of Hydrocarbon-like and Oxygenated Organic Aerosols Based on Aerosol Mass Spectrometry, Environmental Science & Technology, 39, 4938-4952, 2005.

Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra, M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K., DeCarlo, P. F., Salcedo, D., Onasch, T., Jayne, J. T., Miyoshi, T., Shimono, A., Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R. J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenicallyinfluenced Northern Hemisphere midlatitudes, Geophysical Research Letters, 34, L13801, doi:13810.11029/12007GL029979, 2007.

10. Tables

Table 1: Summary of experiments conducted

Experiment ID	Precursor	Structure	k(OH) / k(O_3)[4] / cm^3 $molec^{-1}$ s^{-1}	Experiment Type (no.)	VOC/NO_x Range	RH / % Range
ISOP1 – 10[M,P]	isoprene		9.9×10^{-11} / 1.2×10^{-17}	Photooxidation (10)	1.3 – 20.0	49 – 72
APIN1 – 4[2,3,M,P]	α-pinene		5.3×10^{-11} / 8.4×10^{-17}	Photooxidation (4)	1.3 – 2.0[1]	49 – 73
LIM1 – 6[2,3,M,E,P]	limonene		1.7×10^{-10} / 2.1×10^{-16}	Photooxidation (6)	1.4 – 2.0[1]	50[1] – 82
BCARY1 – 10[2,3,M]	β-caryophyllene		2.0×10^{-10} / 1.2×10^{-14}	Photooxidation (10)	0.6 – 2.0[1]	50[1] – 72
MYRC1 – 2[2,3,M]	myrcene		2.1×10^{-10} / 4.7×10^{-16}	Photooxidation (2)	1.4 – 1.9	52 – 54

Experiment ID	Precursor	Structure	k(OH) / k(O_3)[4] / cm^3 $molec^{-1}$ s^{-1}	Experiment Type (no.)	VOC/NO_x Range	RH / % Range
LINA1 – 2[2,3,M]	linalool		1.6×10^{-10} / 4.5×10^{-16}	Photooxidation (2)	1.4 – 2.6	42 – 47
BIR1 – 2[M]	birch trees	Multiple emissions[5]	Multiple emissions	Mesocosm Photooxidation (2)	5.5 – 5.6	73 – 84
FIG1 – 2[M]	fig trees	Multiple emissions[5]	Multiple emissions	Mesocosm Photooxidation (2)	2.7 – 9.4	65 – 75
TOL1 – 5[E]	toluene		3.7×10^{-12} / -	Photooxidation (5)	1.3 – 11.6	2 – 6

1 = Estimated using known volume of reactants injected
2 = LC-MS/MS filter data available for at least one of these experiments (MAC)
3 = c-TOF-AMS data available for at least one of these experiments (MAC)
4 = From (Atkinson and Arey, 2003b;Sun et al., 2012;Khamaganov and Hites, 2001) and references therein
5 = See Wyche et al., 2014

M = experiments conducted in the MAC

E = experiments conducted in the EUPHORE

P = experiments conducted in the PSISC

Table 2: Key technical features of MAC, EUPHORE and PSISC (Alfarra et al., 2012;Becker, 1996;Bloss et al., 2005;Camredon et al., 2010;Paulsen et al., 2005;Zador et al., 2006).

Chamber	Material	Environment	Size	Light Source	Spectrum
MAC	FEP Teflon	Indoor	18 m^3, 3(H) x 3(L) x 2(W) m	1 x 6 kW Xe arc lamp Bank of halogen lamps	λ range = 290 – 800 nm $j_{NO2} = 6 \times 10^{-4}\ s^{-1}$ (290 – 422 nm)
EUPHORE	FEP Teflon	Outdoor	200 m^3, (hemispherical)	Solar	Solar; 75 % transmission at 290 nm, 85 % transmission > 320 nm $j_{NO2} = \sim 5 - 9 \times 10^{-3}\ s^{-1}$
PSISC	FEP DuPont Tedlar	Indoor	27 m^3, 3(H) x 3(L) x 3(W) m	4 x 4 kW Xenon arc lamps	λ range = 290 – 800 nm $j_{NO2} = 0.12\ min^{-1}$

Table 3: List of certain major product ions integral to the separation of spectra in statistical space, their corresponding tentative assignments and their precursor. See main text, Section 4.2 for further information.

Ion / *m/z*	CIR-TOF-MS		LC-MS/MS		cTOF-AMS	
	Assignment	Precursor	Assignment	Precursor	Assignment	Precursor
41	-	-	-	-	$C_3H_5^+$	All
43	-	-	-	-	CH_3CO^+	All
44	-	-	-	-	CO_2^+	All
71	methyl vinyl ketone + methacrolein	isoprene	-	-	-	-
75	hydroxyl acetone	isoprene	-	-	-	-
83	3-methyl furan	isoprene	-	-	-	-
87	C_4-hydroxycarbonyls / methacrylic acid	Isoprene	-	-	-	-
93	4-vinyl-4-pentenal / 4-hydroxy-4-methyl-5-hexen-1-al	myrcene / linalool	-	-	-	-
95	4-vinyl-4-pentenal / 4-hydroxy-4-methyl-5-hexen-1-al	myrcene / linalool	-	-	-	-
107	piononaldehyde /	α-pinene /	-	-	-	-

	limononaldehyde	limonene				
139	limonaketone	limonene	-	-	-	-
151	piononaldehyde / limononaldehyde	α-pinene / limonene	-	-	-	-
169	piononaldehyde / limononaldehyde	α-pinene / limonene	pinalic-3-acid / ketolimonon-aldehyde + limonalic acid	α-pinene / limonene	-	-
183	-	-	pinonic acid / limononic acid + 7-hydroxylimonon-aldehyde	α-pinene / limonene	-	-
185	-	-	pinic acid / limonic acid	α-pinene / limonene	-	-
199	-	-	$C_9H_{11}O_5$	myrcene	-	-
215	-	-	$C_{10}H_{15}O_5$	myrcene	-	-
235	β-caryophyllene secondary ozonide + isomers	β-caryophyllene	-	-	-	-
227	-	-	$C_{10}H_{11}O_6$	myrcene	-	-
237	β-caryophyllene aldehyde	β-caryophyllene	3-[2,2-dimethyl-4-(1-methylene-4-oxo-butyl)-cyclobutyl]-	β-caryophyllene	-	-

			propanoic acid			
251	-	-	β–caryophyllonic acid	β-caryophyllene	-	-
253	β-caryophyllene secondary ozonide + isomers	β-caryophyllene	-	-	-	-
255	-	-	4-(2-(2-carboxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid	β-caryophyllene	-	-
267	-	-	β-14-hydroxy-caryophyllonic acid + β-10-hydroxy-caryophyllonic acid	β-caryophyllene	-	-
271	-	-	4-(2-(3-hydroperoxy-3-oxopropyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid / 4-(2-(2-carboxy-1-hydroxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid	β-caryophyllene	-	-
321	-	-	$[M-H_2+FA+Na]^-$ $C_{12}H_{14}O_6$ / $C_{13}H_{18}O_5$	myrcene	-	-

Table 4: PLS-DA model classification sensitivity and specificity for the gas-phase biogenic air matrices

Cross Validation	isoprene	cyclic-monoterpene	sesquiterpene	straight-chain-monoterpene	Fig tree	Birch tree
Sensitivity (%)	100.0	100.0	100.0	100.0	100.0	50.0
Specificity (%)	100.0	92.9	100.0	100.0	100.0	91.7

11. Figures

Figure 1: (a) NO_x, O_3, myrcene and 4-vinyl-4-pentenal (primary aldehyde product) and (b) particle mass (not wall loss corrected and assuming $\rho = 1.3$) and size evolution within the MAC during a typical photooxidation experiment.

Figure 2: PCA loadings bi-plot of the second *vs.* first principal components derived from the PCA analysis of the isoprene, cyclic monoterpene ("c-m-terpene" in the legend; α-pinene and limonene), sesquiterpene (β-caryophyllene) and straight chain biogenic ("s-m-terpene" in the legend; myrcene and linalool) chamber data. Classification confidence limits = 95 %. Tentative assignments of major ions, include *m/z* 71 = methyl vinyl ketone and methacrolein, 75 = hydroxy acetone, 83 = e.g. 3-methyl furan, 87 = C_4-hydroxycarbonyls/methacrylic acid, *m/z* 237 = β-caryophyllene aldehyde, 235 and 253 = β-caryophyllene secondary ozonide (and isomers thereof), *m/z* 107, 151 and 169 = piononaldehyde and limononaldehyde, 139 = limonaketone, *m/z* 95 and 93 = 4-vinyl-4-pentenal and 4-hydroxy-4-methyl-5-hexen-1-al. See main text, Section 4.2 and Table 3 for further information. For clarity, the scale has been set to show the bulk of the data, hence precursor parent ions and *m/z* 71 are not shown.

Figure 3: PCA scores plot of the first vs. second principal components derived from the PCA analysis of the mesocosm test set using the PCA model developed from the isoprene, cyclic monoterpene (α-pinene and limonene), sesquiterpene (β-caryophyllene) and straight chain monoterpene (myrcene and linalool) chamber data. Classification confidence limits = 95 %.

Figure 4: PCA scores plot of the first vs. second principal components derived from the PCA analysis of the toluene test set using the PCA model developed from the isoprene, cyclic monoterpene (α-pinene and limonene), sesquiterpene (β-

caryophyllene) and straight chain monoterpene (myrcene and linalool) chamber data. Classification confidence limits = 95 %.

Figure 5: Dendrogram showing the grouping relationship between the various gas-phase matrices of systems examined. Red = isoprene, pink = fig, green = cyclic monoterpenes (α-pinene and limonene), yellow = birch, light blue = straight chain monoterpenes (myrcene and linalool) and dark blue = sesquiterpene (β-caryophyllene).

Figure 6: scores plot of the first three latent variables derived from the PLS-DA model analysis of the isoprene, cyclic monoterpene (α-pinene and limonene), sesquiterpene (β-caryophyllene), straight chain monoterpene (myrcene and linalool), fig and birch chamber data. Classification confidence limits = 95 %.

Figure 7: (a) Loadings bi-plot of the second *vs.* first principal components obtained from the PCA of LC-MS aerosol spectra from a subset of terpene experiments and (b) the corresponding HCA dendrogram. See main text, Section 4.5 and Table 3 for further information, including ion assignments.

Figure 8: (a) Loadings bi-plot of the second *vs.* first principal components obtained from the PCA of AMS aerosol spectra from of a subset of terpene experiments and (b) the corresponding HCA dendrogram. See main text, Section 4.5 and Table 3 for further information, including ion assignments.

Figure 9: Simplified schematic illustrating some of the important mechanistic pathways in the gas-phase oxidation of isoprene, α-pinene, β-caryophyllene and

myrcene, and the associated mass transfer to the particle-phase. Red arrows and text = "high" NO_x pathways, green arrows and text = "low NO_x" pathways, blue arrows and text = ozonolysis reactions, grey arrow and text = speculative, dashed arrows = multiple steps. * = multiple photooxidative routes initiated by reaction with OH (*i.e.* involving the reactants – OH, O_2, NO, HO_2 and/or RO_2), leading to structurally similar products containing different functional groups. *α-pinene mechanism* – X = OH, =O, OOH or ONO_2; Y = CHO or C(O)OH; Z = OH, OOH or ONO_2. *β-caryophyllene mechanism* – X = $CH_2OH(OH)$, $CH_2OH(OOH)$, $CH_2OH(ONO_2)$ or =O. *Myrcene mechanism* – Y = OOH or ONO_2; Z = CHO or C(O)OH. See text, section 5 for references.

Figure 10: PCA loadings bi-plot of the second *vs.* first principal components derived from the PCA analysis of the toluene experiments. Experiments were conducted under low NO_x, high VOC/NO_x ratio (red diamonds), moderate NO_x, medium VOC/NO_x ratio (green squares) and high NO_x, low VOC/NO_x ratio (blue triangle) conditions. For clarity, the scale has been set to show the bulk of the data, hence *m/z* 93 and 85 are not shown.

Figure 11: Results from MCM α-pinene photooxidation simulations. (a) and (b) = basic α-pinene photooxidation; (c) and (d) = spiked injection of α-pinene, continuous HONO input; (e) and (f) = continuous α-pinene and HONO input. Left hand image plots show the evolution of the respective systems over the molecular weight region of interest with time; colour scale = relative abundance (%). Right hand plots = (b) relative abundance of simulated molecular weights during straight α-pinene photooxidation; (d) difference in relative abundance of simulated molecular weights between double injection of α-pinene continuous HONO input and straight α-pinene photooxidation; (f) difference in relative abundance of simulated molecular weights between continuous α-pinene and HONO input and straight α-pinene photooxidation. See text for details.