Molecular composition of biogenic secondary organic aerosols using ultrahigh resolution mass spectrometry: comparing laboratory and field studies

Ivan Kourtchev[1*], Stephen J. Fuller[1], Chiara Giorio[1], Robert M. Healy[2], Eoin Wilson[2], Ian O'Connor[2], John C. Wenger[2], Matthew McLeod[1], Juho Aalto[3,4,5], Taina M. Ruuskanen[5], Willy Maenhaut[6,7], Rod Jones[1], Dean S. Venables[2], John R. Sodeau[2], Markku Kulmala[5], Markus Kalberer[1*]

[1]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK
[2]Department of Chemistry and Environmental Research Institute, University College Cork, Cork, Ireland
[3]Department of Forest Sciences, University of Helsinki, P.O.Box 27, Helsinki, Finland
[4]Hyytiälä Forestry Field Station, Hyytiäläntie 124, Korkeakoski, FI 35500, Finland
[5]Department of Physics, University of Helsinki, P.O. Box 64, Helsinki, Finland
[6]Department of Analytical Chemistry, Ghent University, Krijgslaan 281, S12, BE-9000 Ghent, Belgium
[7]Department of Pharmaceutical Sciences, University of Antwerp, Universiteitsplein 1, BE-2610 Antwerp, Belgium

*Corresponding authors: ink22@cam.ac.uk, mk594@cam.ac.uk

Abstract

Numerous laboratory experiments have been performed in an attempt to mimic atmospheric secondary organic aerosol (SOA) formation. However, it is still unclear how close the aerosol particles generated in laboratory experiments resemble atmospheric SOA with respect to their detailed chemical composition. In this study, we generated SOA in a simulation chamber from the ozonolysis of α-pinene and a biogenic volatile organic compound (BVOC) mixture containing α- and β-pinene, Δ^3-carene, and isoprene. The detailed molecular composition of laboratory-generated SOA was compared with that of background ambient aerosol collected at a boreal forest site (Hyytiälä, Finland) and an urban location (Cork, Ireland) using direct infusion nanoelectrospray ultrahigh resolution mass spectrometry. Kendrick Mass Defect and Van Krevelen approaches were used to identify and compare compound classes and distributions of the detected species. The laboratory-generated SOA contained a distinguishable group of dimers that was not observed in the ambient samples. The presence of dimers was found to be less pronounced in the SOA from the BVOC mixtures when compared to the one component precursor system. The molecular composition of SOA from both the BVOC mixture and *α*-pinene represented the overall composition of the ambient sample from the boreal forest site reasonably well, with 72.3±2.5% (*n*=3) and 69.1 ± 3.0% (*n*=3) common ions, respectively. In contrast, large differences were found between the laboratory-generated BVOC samples and the ambient urban sample. To our knowledge this is the first direct comparison of molecular composition of laboratory-generated SOA from BVOC mixtures and ambient samples.

1. Introduction

Biogenic volatile organic compounds (BVOCs) play an important role in atmospheric chemistry and give rise to secondary organic aerosols (SOA), which have effects on climate (Hallquist et al., 2009) and human health (Pope and Dockery, 2006). SOA is formed within the atmosphere from gaseous precursors and gas-to-particle conversion processes. Laboratory chamber experiments have been performed for decades in an attempt to mimic atmospheric SOA formation. However, it is still unclear how close the aerosol particles generated in laboratory experiments resemble atmospheric SOA with respect to their detailed chemical composition. One of the major challenges is the identification of the organic composition of the SOA, which is composed of thousands of organic compounds (Kanakidou et al., 2005). These compounds generally cover a wide range of polarities, volatilities and masses (Goldstein and Galbally, 2007) and therefore it is difficult to find a single analytical technique for their detailed chemical analysis at the molecular level. Conventional chromatographic methods (gas chromatography (GC) and liquid chromatography (LC)) are not capable of resolving the highly complex mixtures with a wide variety of physico-chemical properties. Moreover, commonly used mass spectrometers, which are often used as detectors following chromatographic separation, do not have sufficient mass-resolving power to distinguish and differentiate all the compounds present in the complex mixture of organic aerosol. Ultra-high resolution mass spectrometers (UHR-MS) (i.e., Fourier transform ion cyclotron resonance MS and Orbitrap MS) have a mass resolution power that is at least one order of magnitude higher (≥100,000) than conventional MS and thus have the potential for solving this problem. Direct infusion electrospray ionisation (ESI)-UHR-MS has been successfully applied for the analysis of both ambient and laboratory-generated SOA and facilitated the characterisation of hundreds of species with individual molecular formulae (Nizkorodov et al., 2011). Despite the high analytical throughput of direct infusion MS, this method is prone to artefacts such as changes in the ionisation efficiency of an analyte due to the presence of ‘matrix’ compounds in the complex organic mixtures (Pöschl, 2005). For instance, sulphates, nitrates and ammonium salts are important constituents of atmospheric aerosols (Pöschl, 2005) and once injected into the ESI source can cause ion suppression, adduct formation and a rapid deterioration of instrument performance (Dettmer et al., 2007). NanoESI-MS, which generally produces smaller droplet sizes and analyte flow in the electrospray (Schmidt et al., 2003), can substantially reduce interference effects from inorganic salts. Moreover, it provides better sensitivity towards a variety of analytes in samples containing relatively high levels of salts (Juraschek et al., 1999, Schmidt et al., 2003) and decreases source contamination (Schmidt et al., 2003) compared to conventional ESI sources.

To date, most laboratory experiments reproducing atmospheric SOA formation have been performed using a single organic precursor (e.g., α- or β-pinene or isoprene) while in the atmosphere a wide range of precursors contribute to SOA, which results in a more complex SOA composition compared to the one-precursor laboratory systems. Although, there are a few studies where oxidation of volatile organic compound (VOC) mixtures were performed, their main goal was to investigate SOA formation, yields (VanReken et al., 2006; Jaoui et al., 2008; Hao et al., 2009; Kiendler-Schar et al., 2009; Mentel et al., 2009; Hao et al., 2011; Hatfield and Huff Hartz, 2011; Waring et al., 2011) and specific products (Jaoui et al., 2003; Amin et al., 2013) rather than detailed molecular composition.

The main objective of this work is to compare the detailed molecular composition of laboratory-generated SOA from oxidation of a single BVOC (α-pinene) and from a mixture of four BVOCs, with samples of ambient aerosol from urban and remote locations using chip-based direct infusion nanoESI-UHR-MS. In a preceding study we examined aerosol samples from the boreal forest site Hyytiälä, Finland, and determined that a dominant fraction of the detected compounds are reaction products of a multi-component mixture of BVOCs (Kourtchev et al., 2013). In the present study we compare the composition of these field samples with SOA generated in chamber experiments from the ozonolysis of α-pinene and of BVOC mixtures containing four species (α- and *β*-pinene, Δ^3-carene, and isoprene) that are most abundant in Hyytiälä's environment. The laboratory experiments were performed under conditions (e.g., relative humidity (RH), aerosol seed, and BVOC ratios) resembling those at the boreal sampling site during the summer 2011 period. To the best of our knowledge this is the first direct comparison of the molecular composition of laboratory-generated SOA from the BVOC mixtures and ambient samples.

2. Materials and methods

2.1 Atmospheric simulation chamber

Experiments were carried out in an atmospheric simulation chamber described in detail elsewhere (Thüner et al., 2004). Briefly, the chamber is a cylinder made of fluorine–ethene–propene (FEP) Teflon® foil with a volume of 3.91 m^3. It was operated at 296 ±2 K using purified air at 0.1-1 mbar above atmospheric pressure. The experiments were performed at

60-68% relative humidity produced from bubbling purified air through heated water. The humidity and temperature were measured using a dew point meter (DRYCAP®DM70 Vaisala). Fans installed at both ends of the chamber were used during the first 5 minutes of the reaction to provide rapid and uniform mixing of the reactants and products. Between experiments the chamber was cleaned by introducing about 1 ppm of ozone into the chamber and flushing with purified air at a flow rate of 0.15 m^3 min^{-1}. The experiments were performed with neutral ammonium sulphate ($(NH_4)_2SO_4$, Sigma Aldrich, 99.99%) seed particles, produced using an atomizer and dried before introduction into the chamber. Aerosol seed particles were subjected to Krypton-85 (Kr-85) charge neutraliser before introduction to the chamber. Seed particle concentrations for each experiment are shown in Table 1. Cyclohexane at a molar concentration 1000 times higher than the BVOC precursors was used to scavenge OH radicals produced from ozonolysis of the reactants. BVOCs (i.e., α-pinene, β-pinene, Δ^3-carene and isoprene) were introduced into the chamber by flowing purified air over known amounts of the compounds in a gently heated Pyrex impinger. The BVOC concentrations are shown in Table 1. Ozone (ca. 200 ppbv) was introduced at the beginning of the reaction over a period of 1 min using an electric discharge generator. Ozone decay was monitored with an automated analyser (Thermo Model 49i). A scanning mobility particle sizer (TSI model 3081) was used to measure particle number-size distributions in the range 11-478 nm (mobility diameter) every 3 min. The particle mass concentrations were calculated assuming a density of 1 g cm^{-3}. Dilution and wall-loss corrections were calculated by considering the particle loss as a first order rate process. The background NO_x concentration measured using a NO_x analyser (Thermo Model 42*i)* was below 2 ppbv for all experiments. The aerosol samples were collected on quartz fibre filters (Pallflex Tissuquartz 2500QAT-UP, 47 mm diameter) 20-30 min after the maximum SOA concentration was observed. Before use, the quartz fibre filters were preheated at 650 ºC for 12 h to remove any possible organic impurities. A charcoal denuder was used in front of the filter pack to remove gas phase species formed during the ozonolysis reaction. The aerosol sampling flow rate was approximately 12 L min^{-1} and the sampling time was 40 min. In addition, ‘blank’ chamber samples were collected by drawing ‘clean’ air from the chamber for 40 min. The filter samples were wrapped in baked aluminium foil and stored at -20 ºC prior to analysis.

2.2 Ambient samples

Biogenic ambient samples were collected at the boreal forest site SMEAR II in Hyytiälä, southern Finland (61°51′N, 24°17′E) as previously described in detail (Kourtchev et al., 2013). The forest around the station is dominated by conifers (mainly Scots pine and Norway spruce) with some deciduous trees, such as aspen and birch, with a tree density of about 2500 ha^{-1}. Detailed descriptions of the site, instrumentation, meteorological data collection and sampling are given elsewhere (Kulmala et al., 2001; Hari and Kulmala, 2005). In total, 10 separate day and night atmospheric aerosol PM_1 samples, each representing 12 hours of sampling, were collected over the period 16 to 25 August 2011.

Anthropogenic ambient samples were collected during 9-17 September, 2011 at the Tivoli Industrial Estate and Docks (TIED), Cork, Ireland (51°54′5 N, 8°24′38 W). A detailed description of the TIED site is given elsewhere (Healy et al., 2009; Hellebust et al., 2010; Kourtchev et al., 2011). The site is located approximately 3 km east of Cork city centre with a population of about 120,000 inhabitants. A shipping berth is located 400–600 m to the southwest, while the main road carrying traffic east out of the city lies 200 m to the north of the site. Residential areas surround the site on all sides except the north and northeast. The vegetation that surrounds the site manly consists of shrubs and some native deciduous trees, which include oak (*Quercus robur L.* and *Quercus petrua*), ash (*Fraxinius excelsior*), birch (*Betula pendula*), sycamore (*Acer pseudoplatanus*), poplar (*Populus tremula*) and beech (*Fagus sylvatica*). It should be noted that there are not many conifer species present around the sampling area and if they do exist, they are likely to be spruce (*Picea sitchensis*). $PM_{2.5}$ aerosol samples were collected on quartz fibre filters (Pallflex Tissuquartz 2500QAT-UP, 150 mm diameter, pre-fired for 24 h at 650°C) using a High Volume (Digitel DHA-80, Switzerland) sampler with a flow rate of 500 L min^{-1}.

2.3 Aerosol sample analysis

All ambient filters were analysed for organic carbon (OC) and elemental carbon (EC) using a thermal-optical transmission (TOT) technique (Birch and Cary, 1996). For each sample, a part of the quartz fibre filter (6–30 cm^2, depending on OC or total aerosol loading for ambient and laboratory samples respectively) was extracted three times with 5 mL of methanol (Optima® grade, Fisher Scientific) under ultrasonic agitation for 30 min. The three extracts were combined, filtered through a Teflon filter (0.2 μm) and reduced by volume to approximately 200 µL under a gentle stream of nitrogen.

The final extracts were analysed using an ultrahigh resolution LTQ Orbitrap Velos mass spectrometer (Thermo Fisher, Bremen, Germany) equipped with a TriVersa Nanomate robotic nanoflow chip-based ESI source (Advion Biosciences, Ithaca NY, USA). The Orbitrap MS instrument was calibrated using an Ultramark 1621 solution (Sigma-Aldrich, UK). The mass accuracy of the instrument was below 1.5 ppm and was routinely checked before the analysis. The instrument mass resolution was 100,000 at *m/z* 400. The negative ionisation mass spectra were collected in three replicates at ranges *m/z* 100-650 and *m/z* 200-900 and processed using Xcalibur 2.1 software (Thermo Scientific). A mixture of camphor, sulphonic acid (20 ng μL^{-1}), glutaric acid (30 ng μL^{-1}) and *cis*-pinonic acid (30 ng μL^{-1}) in methanol and Ultramark 1621 solution were used to optimise the ion transmission settings. The direct infusion nanoESI parameters were as follows: the ionisation voltage and back pressure were set at -1.4 kV and 0.8 psi, respectively.

To assess possible matrix effects caused by inorganic salts on the detection of organic compounds in the direct infusion analysis, the methanolic extracts of the laboratory-generated samples were mixed with 30% aqueous solution of ammonium sulphate (to mimic the ambient concentration ratios in the boreal samples, see discussion below). Control samples were mixed with water in the same proportions. These modified samples were analysed in the same way as the unaltered aerosol extracts.

For the LC/(-)ESI-MS analysis, due to relatively low OC loading of the filter samples, all day and night samples were pooled into one day and one night sample, evaporated to dryness and resuspended in 0.1% formic acid. LC/(-)ESI-MS analysis was performed using an Accela system (Thermo Scientific, San Jose, USA) coupled with LTQ Orbitrap Velos MS and a T3 Atlantis C18 column (3 µm; 2.1 x 150 mm; Waters, Milford, USA). The mobile phases consisted of 0.1% formic acid (v/v) (A) and methanol (B). The applied gradient was as follows: 0–3 min 3% B, 3–25 min from 3% to 50% B (linear), 25–43 min from 50% to 90% B (linear), 43-48 min from 90% to 3% B (linear) and kept for 12 min at 3% B (total run time 60 min). MS spectra were collected in full scan using the lock mass for the deprotonated dimer of formic acid at m/z 91.00368 with a resolution of 100,000 and the mass ranges of *m*/*z* 50-650 and *m*/*z* 150-900. Based on pre-scan information from the full scan MS, a parallel data-dependent collision induced dissociation (CID) multi-stage mass spectrometry (MSn) (n = 1, 2, 3 and 4) was performed on the most intense precursor ion in three scans at the resolution of 30,000.

2.4 Ultrahigh MS resolution data analysis

The ultrahigh resolution mass spectral data interpretation was carried out using a procedure as described elsewhere (Kourtchev et al., 2013). For each direct infusion sample analysis 70-80 mass spectral scans were averaged into one mass spectrum. Molecular assignments were performed using Xcalibur 2.1 software applying the following constraints $^{12}C \leq 100$, $^{13}C \leq 1$,

$^{1}H\leq200$, $^{16}O\leq50$, $^{14}N\leq5$, $^{32}S\leq2$, $^{34}S\leq1$. All mathematically possible elemental formulae with a mass tolerance of ± 5 ppm were calculated. The data filtering was performed using a Mathematica 8.0 (Wolfram Research Inc., UK) code developed in-house that employed several conservative rules and constraints similar to those used in previous studies (Koch et al., 2005; Wozniak et al., 2008; Lin et al., 2012). Only ions with intensities ten times above the noise level were kept for the data analysis. The mass tolerance range for keeping mathematically assigned elemental formulae was set to approximately ± 0.5 ppm and varied within the ±5 ppm tolerance window. This range was determined by establishing the average difference between the theoretical and the experimental mass for nine compounds with known elemental composition determined by LC/MS analyses (Kourtchev et al., 2013). All molecular formulae where $O/C\geq1.2$, $N/C\geq0.5$, $S/C\geq0.2$, $0.3\leq H/C$ and $H/C\geq2.5$ were eliminated with the aim of removing molecules that are not likely to be observed in nature. Moreover, neutral formulae that had either a non-integer or a negative value of the double bond equivalent (DBE) were removed from the list of possible molecules. Double bond equivalents were calculated using Xcalibur 2.1 software. The assigned formulae were additionally checked for the 'nitrogen-rule' and isotopic pattern as described elsewhere (Kourtchev et al., 2013). The background spectra obtained from the procedural blanks were also processed using the rules mentioned above. The formulae lists of the background spectra were subtracted from those of the ambient (or chamber) sample and only formulae with a sample/blank peak intensity ratio ≥10 were retained. All molar ratios, DBE factors and chemical formulae presented in this paper refer to neutral molecules.

2.5 Hierarchical agglomerative cluster analysis

Aerosol samples were classified by hierarchical agglomerative cluster analysis (Lukasová, 1979). The data was organised in a two-way table X_{nm}, where n is the number of samples (six smog chamber samples and two ambient samples) and m is the number of molecules analysed by UHR-MS in the mass range 100-300 Da (451 molecular formulae). The X_{nm} is a binary value indicating the presence/absence of the molecule m in sample n. The cluster analysis was performed using Statistica 10 (StatSoft Inc., Tulsa, OK, USA), based on the unweighted pair-group average linkage method (or average linkage method) and using the percent disagreement (Georgieva et al., 2005) distance measure. The metric used in this study is analogous to the Jaccard's dissimilarity distance measure that is commonly applied for the analyses of binary patterns (Sneath and Sokal, 1973; Anthony et al., 2002; Cordeiro et al., 2003; Kosman and Leonard, 2005). The percent disagreement or simple mismatch metric considers as a match the absence of molecules in all compared spectra. This might give misleading results because two samples could be considered close to each other just because they shared many absences of molecules (Kosman and Leonard, 2005). For this reason, results obtained using the unmodified percent disagreement or different metrics, i.e. Euclidean distance and r-Pearson correlation coefficient, which would have the same drawbacks for binary data, were not considered (Kosman and Leonard, 2005). Therefore, the percent disagreement metric was modified in order to calculate the distances on the basis of the percentage of common ions between the considered samples. The linkage distance between two samples is calculated using the following equation:

$$LD_{ij} = \frac{N - c_{ij}}{N} \times 100$$

where LD_{ij} is the linkage distance between sample i and sample j, N is the total number of ions considered in the cluster analysis (451 ions) and c_{ij} is the number of ions in common between sample i and sample j.

Robustness of the applied technique was evaluated by repeating the cluster analysis using different linkage methods, i.e. single linkage (or nearest neighbour) and complete linkage (farthest neighbour), which gave exactly the same results.

3. Results and Discussion

The BVOC mixture used in the laboratory experiments contained four of the most abundant SOA-precursor monoterpenes (i.e., α-pinene, Δ^3-carene, β-pinene and isoprene) emitted at the boreal forest site in Hyytiälä (Hakola et al., 2003; Aaltonen et al., 2011; Bäck et al., 2012). The emissions of α-pinene and Δ^3-carene were found to be responsible for up to 97% of the total monoterpene proportions in both plant branch emissions from Scots pine trees (a dominant species at Hyytiälä) and the ambient samples from the boreal forest site in Hyytiälä (Bäck et al., 2012). In the present study, SOA ambient samples were collected below the canopy and 5 m above the forest floor; therefore, the BVOC composition is expected to be additionally influenced by emissions from ground vegetation. At the boreal ground floor the monoterpenes were also found to be the most abundant compound group with α-pinene (average 2.975 µg m^{-2} h^{-1}), Δ^3-carene (average 1.305 µg m^{-2} h^{-1}), camphene (average 0.442 µg m^{-2} h^{-1}) and β-pinene (average 0.191 µg m^{-2} h^{-1}) accounting for 90% of the monoterpene fluxes (Aaltonen et al., 2011). Previous studies (Kourtchev et al., 2005, 2008) indicated that SOA from Hyytiälä contained a number of oxidation products of isoprene implying that isoprene certainly plays a role in SOA formation at the boreal site. Therefore, isoprene was added to the BVOC mixture in proportions estimated from the fluxes at the sampling site (Hakola et al., 2003; Aaltonen et al., 2011). Although the total concentrations of the BVOC mixture used in our chamber experiments exceeded those observed at the Finnish site, their molar ratios were kept very close the reported values i.e., α-pinene (0.4), Δ^3-carene (0.3), β-pinene (0.2) and isoprene (0.1).

SOA concentrations and yields from the ozonolysis of α-pinene and BVOC mixtures are shown in Table 1. The average SOA yields (corrected for wall losses) for α-pinene and BVOC mixtures were 0.16±0.01 (*n*=3) and 0.11±0.01 (*n*=3), respectively. The obtained yields for the α-pinene-only experiments are in reasonable agreement with those reported in the literature for similar BVOC concentration ranges (Pathak et al., 2007; Hatfield and Huff Hartz, 2011). Surprisingly, in the present study, SOA yields for the BVOC mixture were significantly lower than for the single BVOC (α-pinene) system. α-pinene and Δ^3-carene accounted for a major fraction (~70%) of the total BVOC mixture and thus are expected to make the major contributions to the SOA mass. Δ^3-carene is reported to have a similar SOA yield to α-pinene (Jonsson et al., 2006, Hatfield and Huff Hartz, 2011) and therefore cannot be responsible for the observed low yield from the precursor mixture. β-pinene and isoprene account for 20% and 10% of the total BVOC mixture, respectively. The ozonolysis of these BVOCs generally results in a lower yield for β-pinene compared to α-pinene (Jonsson et al., 2006) and a very low yield for isoprene (~0.014) (Kleindienst et al., 2007). However, considering that the latter BVOCs account for a small fraction of the reaction mixture their contribution to the total SOA mass is expected to be rather low. It has been shown that the addition of β-pinene to the α-pinene/O_3 system did not affect the SOA yield significantly (Hatfield and Huff Hartz, 2011). Therefore, the possibility that isoprene is responsible for suppressing SOA formation from the precursor mixture cannot be ruled out. Furthermore, as shown in Table 1, ozone was present in all experiments in excess and thus the differences in yield are not expected to be due to the limited availability of oxidants.

3.1 UHR-MS analysis

Representative (-) nanoESI high resolution mass spectra for ambient summer aerosol from the boreal forest site Hyytiälä, Finland, laboratory-generated SOA from ozonolysis of α-pinene, and the BVOC mixture are shown in Figure 1. The molecular composition of the organic aerosol at Hyytiälä is found to be strongly affected by air mass origin. Depending on the sampling day 460-730 molecular formulae were identified in the 10 ambient samples (Kourtchev et al., 2013). The NanoESI mass spectra of the ambient samples is mainly composed of low molecular mass compounds (i.e. peaks below *m/z* 350, Figure 1a) irrespective of the sampling period, which is consistent with previous studies of ambient organic aerosol (OA) from urban (e.g., Riverside, USA (Reemtsma et al., 2006), Cambridge, UK (Rincon et al., 2012), suburban (e.g., Pearl River Delta Region, China (Lin et al., 2012)), and rural (e.g., Mount Werner, north-western Colorado, USA (Mazzoleni et al., 2012)) environments. This is in contrast to laboratory generated SOA from both *α*-pinene (Figure 1c) and the BVOC mixture (Figure 1d) which contain high molecular weight compounds with distinguishable groups of dimers. Similar observations were reported in the literature for laboratory generated SOA from biogenic or anthropogenic BVOCs where UHR mass spectra often contain a large number of oligomers (Reinhardt et al., 2007; Walser et al., 2007; Putman et al., 2012).

Figure 1b shows a mass spectrum containing only those ions that were observed in all 10 Hyytiälä samples (referred to hereafter as 'common ions'). Considering that 'common ions' exclude all species that occurred during the individual days, they are potentially characteristic of locally formed and emitted OA because their presence is independent of the air mass origin. The BVOC mixture samples have a fewer number of peaks in the dimeric region than the *α*-pinene samples. The total number of assigned formulae in the *α*-pinene and BVOC mixture mass spectra were on average 632±84 and 501±54, respectively (where ± describes the variability between three replicate chamber experiments). A higher number of formulae (about 900) were identified from the negative electrospray ultrahigh resolution FT-ICR mass spectra of SOA from *α*-pinene ozonolysis in the previous study of Putman et al., (2012). However, the latter study identified formulae in the range $100<m/z<850$, whereas we only considered ions below *m/z* 650. The number of possible empirical formulae assignments increases significantly with higher masses, especially above 400 Da. Because no common ions >*m/z* 300 are present in the ambient samples, only ions from the monomeric region of the laboratory-generated SOA were used for further comparison with the ambient sample.

In the monomeric region (below *m/z* 300), the number of formulae in SOA from *α*-pinene and the BVOC mixture were comparable, on average 199±29 and 215±17, respectively (Figure 1a and 1b). At first, such a small difference was somewhat puzzling. However, considering that three of the four BVOC precursor compounds (i.e., *α*-pinene, *β*-pinene and Δ^3-carene) used in the mixture are structural isomers, their oxidation with O_3 is expected to yield products with similar elemental composition but different structures, which cannot be separated using the analytical technique employed here. For instance, the mass spectra from both chamber experiments and ambient OA were dominated by an ion at *m/z* 185.0818. While in α-pinene experiments this ion corresponded to *cis*-pinic acid, in the BVOC mixture experiments and Hyytiälä ambient samples this ion was related to three (i.e. *cis*-pinic acid, homoterpenylic acid, and *cis*-caric acid) and five (i.e. *cis*-pinic acid, homoterpenylic acid, limonic acid, ketolimononic acid and *cis*-caric acid) different compounds, respectively. The separation and identification of these compounds was achieved using LC/MS analysis. The mass spectral differences observed in the monomeric region of SOA produced from the single precursor and BVOC mixture were mainly associated with variation of the ion intensities caused by differences in the number of products with the same mass in the single component and BVOC mixture experiments as described above.

The ionisation of organic compounds can be affected by the presence of inorganic salts in the analyte solutions, potentially leading to a decrease in MS signal intensity when using direct infusion mass spectrometry methods. Thus, we tested whether the presence of atmospherically abundant salts (e.g., ammonium sulfate) in our filter extracts could cause such a matrix effect and whether this could be responsible for the lack of dimers observed in the ambient samples. Laboratory generated samples were spiked with ammonium sulfate at atmospherically realistic proportions (30% of the total aerosol mass). The addition of salts suppressed the intensities of all ions in the entire mass range but did not selectively decrease the intensity of ions in the dimeric region (Figure S1). However, due to competitive ionisation of analytes in the ESI direct infusion analysis of the aerosol samples that are known to have a very complex matrix, the ion intensities do not directly reflect the concentration of the molecules in the sample. Therefore, signal intensities should be interpreted with caution and thus were not considered for the mass spectral comparison in this study. In contrast, LC/MS which is a quantitative technique showed significant difference in the abundances of peaks associated with higher-molecular weight (HMW) compounds between ambient and laboratory generated samples (Figure 2). While a number of HMW species associated with *m/z* 337.1652, 357.1552, 367.1762, and 377.1454 were observed in the chromatogram from laboratory generated SOA (Figure 2a), only one of these species (*m/z* 357.1552) was detected in the ambient samples with intensity just above the chromatographic noise (Figure 2b). It should be noted that a chromatographic peak associated with *m/z* 357.1552 was the most dominant species along with *cis*-pinic acid (*m/z* 185.0818) in all laboratory samples. Further MS/MS fragmentation of the major ion at *m/z* 357.1552 resulted in two abundant product ions at *m/z* 185.0818 and 171.0662. Similar observations were reported by Yasmeen et al. (2010), who suggested that the HMW compound adt *m/z* 357 is a possible esterification product of *cis*-pinic and diaterpenylic acid. Both of these acids were found to be very abundant in our ambient and laboratory generated samples; however, as outlined above, their dimer was only present in the latter samples (Figure 2). It is worth mentioning that the HMW compound at *m/z* 357 has been previously identified in SOA produced from the ozonolysis of α-pinene and attributed to a hydroxyperoxyhemiacetal (Müller et al., 2009) and pinyl-diaterpenylic ester (Kristensen et al., 2013)

These results rule out the possibility that the observed direct infusion nanoESI mass spectral differences in the ambient and laboratory generated samples are either due to matrix or methodological artifacts. We can thus conclude that the dimer concentration in the boreal forest OA is negligible compared to the laboratory SOA.

The Van Krevelen diagram, in which the H/C ratio is plotted as a function of the O/C ratio for each formula in a sample, is often used to describe the evolution of organic mixtures. Moreover, VK diagrams can also be used to visualise the differences in the elemental composition of different samples. Figure 3 shows an overlaid VK diagram for SOA from (a) α-pinene, the BVOC mixture and a boreal forest sample from Hyytiälä and (b) α-pinene, the BVOC mixture and a sample from the TIED site, which is heavily influenced by anthropogenic emissions. As indicated above, the elemental composition from the boreal forest site only included ‘common ions’ because they are potentially characteristic for locally emitted OA as their presence is independent of the air mass origin. The elemental ratios from the TIED site included ‘common ions’ from 3-4 September 2011, associated with westerly air masses. The composition of the latter samples is discussed in detail in a separate article (Kourtchev et al., prep). It can be seen from Figure 3, that the distribution of elemental ratios of laboratory-generated SOA from α-pinene is very similar to that of the BVOC mixture. Moreover, the elemental distribution of laboratory SOA generated from a single precursor and a mixture of BVOCs represent fairly well that of the ambient SOA from Hyytiälä, except that the latter sample contained an additional cluster of molecules as displayed at the upper left part of the diagram. In general, this region is associated with the most reduced/saturated species (Lin et al., 2012) and could therefore possibly be fatty acids emitted from a local biogenic source (Kourtchev et al., 2013). In contrast, the VK diagrams of the laboratory-generated SOA were very different from that of the anthropogenic aerosol from the TIED site, which contained a large cluster of ions with low H/C ($\leq$1.0) and O/C ($\leq$0.5) ratios, possibly corresponding to oxidised aromatic hydrocarbons (Mazzoleni et al., 2012). These differences were also apparent when the data was expressed as DBE versus mass-to-charge-ratio (*m/z*) (Figure 4). The samples from α-pinene and the BVOC mixtures had very similar DBE distribution with values in the range 1 to 7. A small number of species observed in the laboratory samples with DBE values of 5-7 were possibly associated with dimers that were formed through accretion reactions (Putman et al., 2012). The DBE distribution of molecules from the Hyytiälä samples clearly resembled those of the α-pinene and BVOC mixtures, except that the ambient sample contained an additional cluster of ions with DBE 0-1. As determined by MS/MS analysis they are attributed to unsaturated and saturated fatty acids (Kourtchev et al., 2013). The DBE plot for OA from TIED was very different from the rest of the compared samples and contained an additional large cluster of molecules with DBE between 7 and 13 (Figure 4b), once more indicating the presence of oxidised aromatic species. Aromatic compounds are typically associated with anthropogenic sources (Henze et al., 2008) whereas aliphatic compounds can be of both anthropogenic and biogenic origin.

The average O/C and H/C ratios for SOA from α-pinene (0.55 and 1.46, respectively) and the BVOC mixture (0.58 and 1.40) were fairly comparable to the ratios for OA from Hyytiälä (0.52 and 1.48) (Kourtchev et al., 2013) but higher than those from TIED (0.36 and 1.1). The H/C value for laboratory-generated SOA indicated that the identified SOA molecules are of aliphatic and alicyclic nature (Putman et al., 2012). The elemental O/C ratios found in this study are within the range obtained for SOA generated in laboratory experiments from a number of BVOCs, e.g., α-pinene/O_3 (0.42-0.55) (Putman et al., 2012), and photo-oxidation of isoprene under low-NO_x conditions (0.54) (Nguyen et al., 2011). These values are slightly higher than the average values (0.43-0.46) reported for the limonene/O_3 sytem (Bateman et al., 2009).

It has been demonstrated that O/C ratio, as measured by the Aerodyne High Resolution Time-of-Flight Aerosol Mass Spectrometer (HR-ToF-AMS), is positively correlated with the hygroscopic parameter of the organic fraction (Wu et al., 2013), which in turn is related to the cloud condensation nucleus (CCN) activity of aerosol particles (Petters and Kreidenweis, 2007). Therefore, considering that the O/C ratio for SOA generated from the α-pinene-only system is very similar to that from the BVOC mixture and OA from Hyytiälä, we suggest that the simplified BVOC system can possibly be used for parameterisation of OA at the boreal site.

O/C ratios may not accurately describe the degree of oxidation of organics, because other non-oxidative processes (e.g., hydration and dehydration) can also affect these parameters (Kroll et al., 2011). Carbon oxidation state (OS_C), on the other hand, may change differently upon oxidation, but the average OS_C will increase. Therefore, OS_C could be a useful metric for the degree of oxidation of organic species in the atmosphere and can serve as a key variable to describe organic mixtures. OS_C is shown to be strongly linked to aerosol volatility and thus is a useful parameter for the classification of SOA (Hao et al., 2011). Carbon oxidation state can be calculated from the following equation:

$$OS_C = -\sum_i OS_i \frac{n_i}{n_C}$$

where OS_i is the oxidation state associated with element *i* and n_i/n_C is the molar ratio of element *i* to carbon (Kroll et al., 2011). Figure 5 shows the overlaid carbon oxidation state versus the number of carbon atoms for molecules from laboratory-generated SOA and the ambient boreal samples. The OS_C distribution of laboratory-generated SOA generally resembles that of the ambient samples ranging from -1.7 to 1 and only a few species have an oxidation state greater than +1. Interestingly, molecules with OS_C greater than +1 were only observed in SOA from the BVOC mixture experiments and ambient aerosol from Hyytiälä. Such compounds are expected to contain several carbonyl groups. However, literature data suggests that the average oxidation state of organic aerosol rarely exceeds this value because species with several carbonyl groups are highly unstable and will rapidly decompose to smaller molecules (Kroll et al., 2011). Considering that the studied BVOC mixture mainly contained monoterpenes, which are structural isomers and the fact that highly oxidised molecules were not observed in the SOA generated from the ozonolysis of α-pinene, it is likely that species with $OSc>+1$ were produced from the ozonolysis of isoprene. Moreover, in the BVOC mixture experiments cross reactions between radicals and oxidation products of the different BVOCs are expected to occur which may lead to formation of a complex range of species. These highly oxidised species are worthy of further investigation. It should be mentioned that highly oxidised multifunctional molecules have been previously observed in both ambient air (Hyytiälä) and in chamber experiments of α- and β-pinene ozonolysis (Ehn et al., 2012).

The majority of the species exhibited OS_C values between – 1 and +1 with 15 or less carbon atoms, suggesting that they are semi- and low-volatile organic compounds corresponding to 'fresh' and 'aged' SOA produced by multistep oxidation reactions (Jimenez et al., 2009, Kroll et al., 2011). Compared to chamber samples, the Hyytiälä samples additionally contained ions with $OSc<-1$ and more than 7 carbon atoms which is characteristic of primary biomass burning aerosol (Kourtchev et al., 2013).

3.2 Comparison using statistical tools

Figure 6 shows the fraction of molecular formulae below 300 Da found in both the laboratory-generated SOA and the ambient samples relative to the total number of formulae in the ambient samples. Evidently, the molecular composition of SOA from both the BVOC mixture and α-pinene represented the overall composition of the ambient sample from the boreal forest site reasonably well, with 72.3±2.5% (n=3) and 69.1 ± 3.0% (n=3) common ions, respectively. Although, the BVOC mixture resulted in a slightly higher number of common formulae than that from the boreal forest aerosol compared to the pure α-pinene-SOA, the difference in the mean values among the treatment groups is not large enough to exclude the possibility that the difference is due to random sampling variability; according to ANOVA test, the difference was not statistically significant ($p = 0.348$). In contrast, the molecular composition of laboratory-generated SOA was substantially different from that of the anthropogenically affected TIED site. The fraction of common molecular formulae from α-pinene and the BVOC mixture relative to the total number of ions from the TIED sample was only 16.1±1.7% and 16.9±1.2%, respectively, indicating the very different sources of organic compounds in these samples. The inverse comparison of the fraction of common formulae relative to the total number of formulae in the laboratory generated samples in the monomeric region indicated that laboratory generated SOA contained 20-25% formulae that were not observed in the boreal samples. We suggest that these molecules are first generation products that are possibly oxidised with time in the atmosphere resulting in aged oxidation products.

Laboratory-generated and ambient samples were also compared by hierarchical cluster analysis (HCA) that divides samples into groups (clusters) of similar molecular composition. HCA separated the samples into three clusters (Figure 7): (1) α-pinene (replicates from three different experiments); (2) BVOC mixture (three replicates) together with the common ions of the ambient samples from Hyytiälä; and (3) common ions of the ambient samples from TIED. The branches in the tree diagram (dendrogram) represent the average distance between the connected samples. It is evident from the dendrogram that all replicate samples from the α-pinene and the BVOC mixture experiments cluster together, implying very good reproducibility of the applied technique (i.e., smog chamber experimental and MS analysis) to separate two experimental conditions relative to each other. Although the α-pinene data is separated from the BVOC mixture and Hyytiälä cluster, the linkage distance is not large enough to conclude that their chemical composition is very different. On the other hand, the data from TIED was classified into a separate cluster confirming that its molecular composition is very different from the rest of the samples. The results from HCA clearly support the findings obtained from the statistical analysis and other visualisation methods (Van Krevelen diagrams, carbon oxidation state, DBE).

4. Conclusions

The detailed molecular composition of background ambient aerosol from a boreal forest site (Hyytiälä, Finland), an urban location (Cork, Ireland), laboratory-generated SOA from α-pinene and a mixture of four BVOCs were compared using nanoESI UHRMS. Our results demonstrate that the molecular composition of SOA in the monomeric mass range up to *m/z* 300 from both the ozonolysis of the BVOC mixture and α-pinene represented the overall composition of the ambient sample from the boreal forest site fairly well, with 72.3±2.5% (n=3) and 69.1 ± 3.0% (n=3) common ions, respectively. Other atmospheric oxidants (e.g., OH radicals and NO_x) will certainly influence the composition of SOA and their reaction products are likely to explain some of the remaining molecules that were not observed in our laboratory generated SOA.

The elemental (O/C and H/C) ratios of SOA from the α-pinene-only system were very similar to those from the BVOC mixtures and ambient aerosol from boreal forest. Considering that the O/C ratio is positively correlated with hygroscopicity of the organic fraction, the simplified α-pinene-only system can potentially be useful for parameterisation of boreal OA. A specific class of CHO compounds identified as fatty acids was present exclusively in the ambient samples suggesting that the composition at the boreal forest OA is also influenced by primary emissions. In contrast, the overall molecular composition of the urban samples is dominated by a high number of oxidised aromatic hydrocarbons and is very different from the boreal and laboratory-generated OA. The presence of dimers observed exclusively in the laboratory samples requires further investigation. Thus it appears that while the laboratory experiments performed here and in other studies do a fairly good job of simulating biogenic SOA formation (as indicated by the high level of agreement between the laboratory SOA and the boreal forest samples in the monomeric mass region of the mass spectra), they do not completely replicate the processes occurring in the ambient atmosphere.

Acknowledgements

Research at the University of Cambridge was supported by a Marie Curie Intra-European fellowship (project # 254319) and the European Research Council (ERC starting grant 279405). The work at University College Cork was funded by the European Commission (project EUROCHAMP-2, contract number 228335). The work at University of Helsinki was supported by Academy of Finland Center of Excellence program (project number 1118615). The authors would like to acknowledge the use of NAME, developed and provided by the Atmospheric Dispersion Group, UK Met Office.

Supporting Information Available

One supporting figure (Figures S1).

References:

Aaltonen, H., Pumpanen, J., Pihlatie, M., Hakola, H., Hellén, H., Kulmala, L., Vesala, T., and Bäck, J. Boreal pine forest floor biogenic volatile organic compound emissions peak in early autumn, Agric. Forest Meteorol., 151, 682−691, 2011.

Amin, H. S., Hatfield, M. L., and Huff Hartz, K. E.: Characterization of secondary organic aerosol generated from ozonolysis of α-pinene mixtures, Atmos. Environ., 67, 323-330, 2013.

Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P.: The origin of cultivated *Coffea arabica L.* varieties revealed by AFLP and SSR markers, Theor. Appl. Genet., 104, 894–900, 2002.

Bäck, J., Aalto, J., Henriksson, M., Hakola, H., He, Q., and Boy, M.: Chemodiversity of a Scots pine stand and implications for terpene air concentrations, Biogeosciences, 9, 689–702, 2012.

Bateman, A. P., Nizkorodov, S. A., Laskin, J., and Laskin, A.: Time-resolved molecular characterization of limonene/ozone aerosol using high-resolution electrospray ionization mass spectrometry, Phys. Chem. Chem. Phys., 11, 7931-7942, 2009.

Birch, M. E. and Cary, R. A.: Elemental carbon-based method for monitoring occupational exposure to particulate diesel exhaust, Aerosol Sci. Technol., 25, 221–241, 1996.

Cordeiro, G. M., Pan, Y. B., and Henry, R. J.: Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm, Plant Sci., 165, 181–189, 2003.

Dettmer, K., Aronov, P. A., and Hammock, B. D.: Mass spectrometry-based metabolomics, Mass Spectrom. Rev., 26, 51–78, 2007.

Ehn, M., Kleist, E., Junninen, H., Petäjä, T., Lönn, G., Schobesberger, S., Dal Maso, M., Trimborn, A., Kulmala, M., Worsnop, D.R., Wahner, A., Wildt, J., and Mentel, Th. F.:

Gas phase formation of extremely oxidized pinene reaction products in chamber and ambient air, Atmos. Chem. Phys., 12, 5113-5127, 2012.

Georgieva, E., Handjieva, N., Popov, S., and Evstatieva, L.: Comparative analysis of the volatiles from flowers and leaves of three *Gentiana* species, Biochem. Syst. Ecol., 33, 938-947, 2005.

Goldstein, A. H. and Galbally, I. E.: Known and unexplored organic constituents in the Earth’s atmosphere, Environ. Sci. Technol., 41(5), 1514–1521, 2007.

Hakola, H., Tarvainen, V., Laurila, T., Hiltunen, V., Hellen, H., and Keronen, P.: Seasonal variation of VOC concentrations above a boreal coniferous forest, Atmos. Environ., 37, 1623−1634, 2003.

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, Th. F., Monod, A., Prévôt, 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, Atmos. Chem. Phys., 9, 5155-5236, 2009.

Hatfield, M. L. and Huff Hartz, K. E. H.: Secondary organic aerosol from biogenic volatile organic compound mixtures source, Atmos. Environ., 45, 2211- 2219, 2011.

Hao, L. Q., Yli-Pirila, P., Tiitta, P., Romakkaniemi, S., Vaattovaara, P., Kajos, M. K., Rinne, J., Heijari, J., Kortelainen, A., Miettinen, P., Kroll, J. H., Holopainen, J. K., Smith, J. N., Joutsensaari, J., Kulmala, M., Worsnop, D. R., and Laaksonen, A.: New particle formation from the oxidation of direct emissions of pine seedlings, Atmos. Chem. Phys., 9, 8121-8137, 2009.

Hao, L. Q., Romakkaniemi, S., Yli-Pirilä, P., Joutsensaari, J., Kortelainen, A., Kroll, J. H., Miettinen, P., Vaattovaara, P., Tiitta, P., Jaatinen, A., Kajos, M. K., Holopainen, J. K., Heijari, J., Rinne, J., Kulmala, M., Worsnop, D. R., Smith, J. N., and Laaksonen, A.: Mass yields of secondary organic aerosols from the oxidation of α-pinene and real plant emissions, Atmos. Chem. Phys., 11, 1367–1378, 2011.

Hari, P., and Kulmala, M.: Station for measuring ecosystem− atmosphere relations (SMEAR II), Boreal Env. Res., 10, 315−322, 2005.

Healy, R. M., O'Connor, I. P., Hellebust, S., Allanic, A., Sodeau, J. R., and Wenger, J. C.: Characterisation of single particles from in-port ship emissions, Atmos. Environ., 43, 6408–6414, 2009.

Hellebust, S., Allanic, A., O'Connor, I. P., Wenger, J. C., and Sodeau, J. R.: The use of real-time monitoring data to evaluate major sources of airborne particulate matter, Atmos. Environ., 44, 1116–1125, 2010.

Henze, D. K., Seinfeld, J. H., Ng, N. L., Kroll, J. H., Fu, T.-M., Jacob, D. J., and Heald, C. L.: Global modeling of secondary organic aerosol formation from aromatic hydrocarbons: high- vs. low-yield pathways, Atmos. Chem. Phys., 8, 2405-2420, 2008.

Jaoui, M. and Kamens, R. M.: Gaseous and particulate oxidation products analysis of a mixture of alpha-pinene plus beta-pinene/O_3/air in the absence of light and alpha-pinene plus beta-pinene/NOx/air in the presence of natural sunlight, J Atmos. Chem., 44, 259–297, 2003.

Jaoui, M., Edney, E. O., Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., Surratt, J. D., and Seinfeld, J. H.: Formation of secondary organic aerosol from irradiated alpha-pinene/toluene/NOx mixtures and the effect of isoprene and sulfur dioxide, J Geophys. Res., 113, D09303, doi:10.1029/2007JD009426, 2008.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, A. S. H., Zhang, Q., Kroll, J.H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, A. C., Docherty, K. S., Ulbrich, I. M., Grieshop, A. P., Robinson, A. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V.A., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. A., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, A. M., Williams, L. R., Wood, E. C., Middlebrook, A. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere, Science, 326, 1525–1529, 2009.

Jonsson, A. M., Hallquist, M., and Ljungstom, E.: Impact of humidity on the ozone initiated oxidation of limonene, Δ3-carene, and α-pinene, Environ. Sci. Technol., 40, 188-194, 2006.

Juraschek, R., Dülcks, T., and Karas, M.: Nanoelectrospray—more than just a minimized-flow electrospray ionization source, J. Am. Soc. Mass. Spectr., 10(4), 300–308, 1999.

Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Dentener, F. J., Facchini, M. C., Van Dingenen, R, Ervens, B., Nenes, A., and Nielsen, C. J.: Organic aerosol and global climate modelling: a review, Atmos. Chem. Phys., 5, 1053–1123, 2005.

Kiendler-Scharr, A., Wildt, J., Dal Maso, M., Hohaus, E., Mentel, T. F., Tillmann, R.,Uerlings, R., Schurr, U., and Wahner, A.: New particle formation in forests inhibited by isoprene emissions, Nature, 461, 381-384, 2009.

Kleindienst, T. E., Lewandowski, M., Offenberg, J. H., Jaoui, M., and Edney, E. O.: Ozone-isoprene reaction: Re-examination of the formation of secondary organic aerosol, Geophys. Res. Lett., 34, L01805, doi:10.1029/2006GL027485, 2007.

Koch, B. P., Witt, M., Engbrodt, R., Dittmar, T., and Kattner, G.: Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometry, Geochim. Cosmochim. Acta, 69, 3299–3308, 2005.

Kosman, E. and Leonard, K. J.: Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species, Mol. Ecol., 14, 415-424, 2005.

Kourtchev, I., Ruuskanen, T., Maenhaut, W., Kulmala, M., and Claeys, M.: Observation of 2-methyltetrols and related photooxidation products of isoprene in boreal forest aerosols from Hyytiälä, Finland, Atmos. Chem. Phys., 5, 2761−2770, 2005.

Kourtchev, I., Ruuskanen, T. M., Keronen, P., Sogacheva, L., Dal Maso, M., Reissell, A., Chi, X., Vermeylen, R., Kulmala, M., Maenhaut, W., and Claeys, M.: Determination of isoprene and α-/β-pinene oxidation products in boreal forest aerosols from Hyytiälä, Finland: Diel variations and possible link with particle formation events, Plant Biol., 10, 138– 149, 2008.

Kourtchev, I., Hellebust, S., Bell, J. M., O'Connor, I. P., Healy, R. M., Allanic, A., Healy, D., Wenger, J. C., and Sodeau, J. R.: The use of polar organic compounds to estimate the contribution of domestic solid fuel combustion and biogenic sources to ambient levels of organic carbon and PM2.5 in Cork Harbour, Ireland, Sci. Tot. Environ., 409, 2143-2155, 2011.

Kourtchev, I., Fuller, S., Aalto, J., Ruuskanen, T. M., McLeod, M. W., Maenhaut, W., Jones, R., Kulmala, M., and Kalberer, M.: Molecular composition of boreal forest aerosol from Hyytiälä, Finland, using ultrahigh resolution mass spectrometry, Environ. Sci. Technol., 47, 4069-4079, 2013.

Kristensen, K., Enggrob, K. L., King, S. M., Worton, D. R., Platt, S. M., Mortensen, R., Rosenoern, T., Surratt, J. D., Bilde, M., Goldstein, A. H., and Glasius, M.: Formation and occurrence of dimer esters of pinene oxidation products in atmospheric aerosols, Atmos. Chem. Phys., 2013.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nat. Chem., 3, 133–139, 2011.

Kulmala, M., Hameri, K., Aalto, P. P., Mäkelä, J. M., Pirjola, L., Nilsson, E. D., Buzorius, G., Rannik, U., Maso, M., Seidl, W., Hoffman, T., Janson, R., Hansson, H.-C., Viisanen, Y., Laaksonen, A., and O’Dowd, C.: Overview of the international project on biogenic aerosol formation in the boreal forest (BIOFOR), Tellus, 53B, 324−343, 2001.

Lin, P., Rincón, A. G., Kalberer, M., and Yu, J. Z.: Elemental composition of HULIS in the Pearl River Delta Region, China: Results inferred from positive and negative electrospray high resolution mass spectrometric data, Environ. Sci. Technol., 46 (14), 7454-7462, 2012.

Lukasová, A.: Hierarchical agglomerative clustering procedure, Patt. Recogn., 11, 365-381, 1979.

Mazzoleni, L. R., Saranjampour, P., Dalbec, M. M., Samburova, V., Hallar, A. G., Zielinska, B., Lowenthal, D. H., and Kohl. S.: Identification of water-soluble organic carbon in non-urban aerosols using ultrahigh-resolution FT-ICR mass spectrometry: organic anions, Environ. Chem., 9, 285–297, 2012.

Mentel, T. F., Wildt, J., Kiendler-Scharr, A., Kleist, E., Tillmann, R., Dal Maso, M., Fisseha, R., Hohaus, T., Spahn, H., Uerlings, R., Wegener, R., Griffiths, P. T., Dinar, E., Rudich, Y., and Wahner, A.: Photochemical production of aerosols from real plant emissions, Atmos. Chem. Phys., 9, 4387-4406, 2009.

Müller, L., Reinnig, M. C., Hayen, H., and Hoffmann, T.: Characterization of oligomeric compounds in secondary organic aerosol using liquid chromatography coupled to electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Rapid Commun. Mass Spectrom., 23, 971-979, 2009.

Nguyen, T. B., Roach, P. J., Laskin, J., Laskin, A., and Nizkorodov, S.A.: Effect of humidity on the composition of isoprene photooxidation secondary organic aerosol, Atmos. Chem. Phys., 11, 6931-6944, 2011.

Nizkorodov, S. A., Laskin, J., and Laskin, A.: Molecular chemistry of organic aerosols through the application of high resolution mass spectrometry, Phys. Chem. Chem. Phys., 13, 3612–3629, 2011.

Pathak, R., Stanier, C., Donahue, N. M., and Pandis, S. N.: Ozonolysis of α-pinene at atmospherically relevant concentrations: temperature dependence of aerosol mass fractions, J. Geophys. Res., 112, D03201, doi:10.1029/2006JD007436, 2007.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961-1971, 2007.

Pope, C. A. III and Dockery, D. W.: Health effects of fine particulate air pollution: lines that connect, J Air Waste Manage. Assoc., 56, 709–742, 2006.

Pöschl, U.: Atmospheric aerosols: composition, transformation, climate and health effects, Angew. Chem. Int. Ed. Engl., 44, 7520–7540, 2005.

Putman, A. L., Offenberg, J. H., Fisseha, R., Kundu, S., Rahn, T. A., and Mazzoleni, L. R.: Ultrahigh-resolution FT-ICR mass spectrometry characterization of alpha-pinene ozonolysis SOA, Atmos. Environ., 46, 164–172, 2012.

Reemtsma, T., These, A., Venkatachari, P., Xia, X. Y., Hopke, P. K., Springer, A., and Linscheid, M.: Identification of fulvic acids and sulphated and nitrated analogues in atmospheric aerosol by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Anal. Chem., 78, 8299-8304, 2006.

Reinhardt, A., Emmenegger, C., Gerrits, B., Panse, C., Dommen, J., Baltensperger, U., Zenobi, R., and Kalberer, M.: Ultrahigh mass resolution and accurate mass measurements as a tool to characterize oligomers in secondary organic aerosols, Anal. Chem., 79, 4074–4082, 2007.

Rincón, A. G., Calvo, A. I., Dietzel, M., Kalberer, M.: Seasonal differences of urban organic aerosol composition – an ultra-high resolution mass spectrometry study, Environ. Chem.**,** 9, 298-319, 2012.

Schmidt, A., Karas, M., and Dulcks, T.: Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? J. Am. Soc. Mass Spectrom., 14, 492-500, 2003.

Sneath, P. A. and Sokal, R. R. (Eds.): Numerical Taxonomy, W.H. Freeman Co, San Francisco, 1973.

Thüner, L. P., Bardini, P., Rea, G. J., and Wenger, J. C.: Kinetics of the gas-phase reactions of OH and NO_3 radicals with dimethylphenols, J. Phys. Chem. A, 108, 11019-11025, 2004.

VanReken, T. M., Greenberg, J.P., Harley, P. C., Guenther, A. B., and Smith, J. N.: Direct measurement of particle formation and growth from the oxidation of biogenic emissions, Atmos. Chem. Phys., 6, 4403-4413, 2006.

Walser, M. L., Park, J., Gomez, A. L., Russell, A. R., and Nizkorodov, S. A.: Photochemical aging of secondary organic aerosol particles generated from the oxidation of d-limonene, J. Phys. Chem. A**,** 111, 1907–1913, 2007.

Waring, M. S., Wells, J. R., and Siegel, J. A.: Secondary organic aerosol formation from ozone reactions with single terpenoids and terpenoid mixtures, Atmos. Environ., 45, 4235–4242, 2011.

Wozniak, A. S., Bauer, J. E., Sleighter, R. L., Dickhut, R. M., and Hatcher, P. G.: Molecular characterization of aerosol-derived water soluble organic carbon using ultrahigh resolution electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, Atmos. Chem. Phys., 8, 5099-5111, 2008.

Wu, Z. J., Poulain, L., Henning, S., Dieckmann, K., Birmili, W., Merkel, M., van Pinxteren, D., Spindler, G., Müller, K., Stratmann, F., Herrmann, H., and Wiedensohler, A.: Relating particle hygroscopicity and CCN activity to chemical composition during the HCCT-2010 field campaign, Atmos. Chem. Phys., 13, 7983-7996, 2013.

Yasmeen, F., Vermeylen, R., Szmigielski, R., Iinuma, Y., Böge, O., Herrmann, H., Maenhaut, W., and Claeys, M.: Terpenylic acid and related compounds: precursors for dimers in secondary organic aerosol from the ozonolysis of α- and β-pinene, Atmos. Chem. Phys., 10, 9383-9392, doi:10.5194/acp-10-9383-2010, 2010.

Table and Figures:

Table 1. Summary of α-pinene and BVOC mixture ozonolysis experiments.

Experiment	Total BVOC conc., ppb	Ozone conc., ppb	RH, %	Seed conc., µg m^{-3}	ΔM, µg m^{-3}	SOA yield (Y)$^{(a,b)}$
α-pinene	49.5	200	60	3.1	42.3	0.15
α-pinene	50.5	200	64	3.0	50.0	0.16
α-pinene	55.2	200	62	4.3	55.9	0.16
BVOC mixture	62.9	200	63	5.2	34.8	0.11
BVOC mixture	67.7	200	68	3.8	43.6	0.12
BVOC mixture	62.9	200	65	3.6	41.8	0.12

(a)Yields were calculated from the equation $Y = \Delta M/\Delta HC$, where ΔM (μg m^{-3}) is the particle mass loading and ΔHC (μg m^{-3}) is the reacted concentration of BVOCs; (b) The data is wall loss corrected.

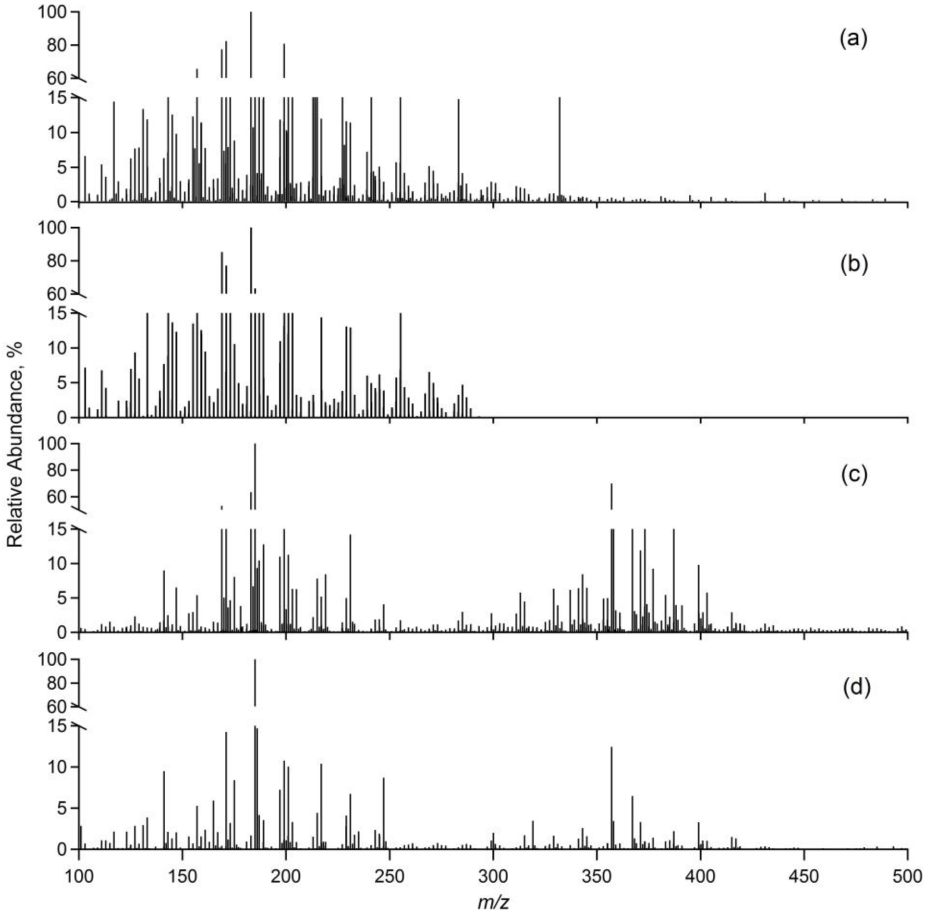

Figure 1. Direct infusion negative-nanoESI-Orbitrap mass spectra obtained for representative OA samples from (a) boreal forest, Hyytiälä, Finland collected on 17 August (night-time), (b) boreal forest, Hyytiälä showing only ‘common’ ions that are present in all 10 samples (see text for detailed explanation), (c) α-pinene/O_3 experiments and (d) BVOC mixture/O_3 experiments. For clarity, the most intense peak at *m/z* 212.0751 has been removed and mass spectra (a) and (b) have been normalised to the second most intense ion at *m/z* 185.0818. The mass spectrum for SOA from the BVOC mixture/O_3 has been scaled up to display all ions that were suppressed by the most intense ion at *m/z* 185.0818.

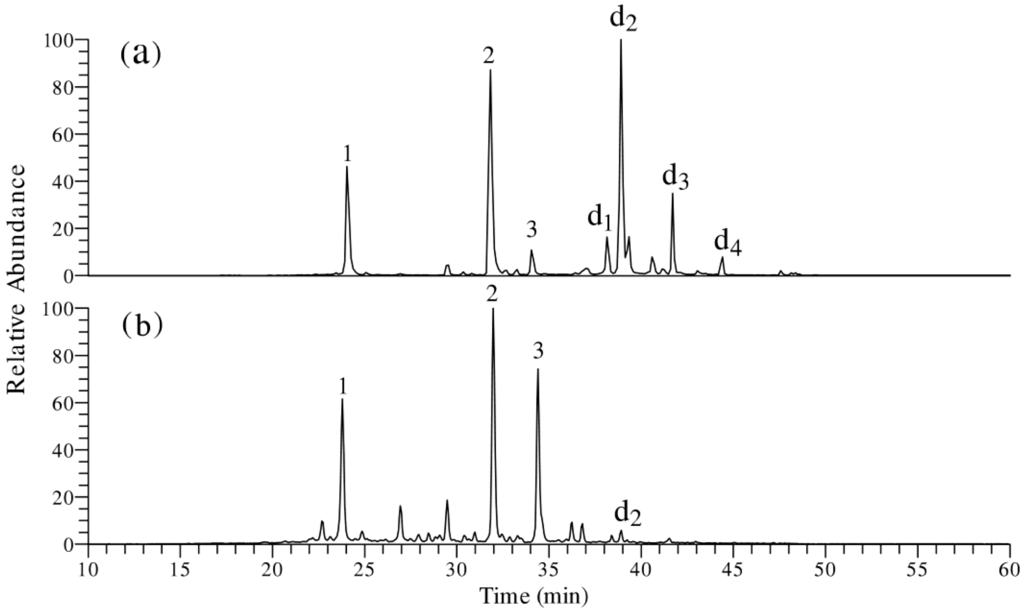

Figure 2. LC/(−)ESI-MS extracted ion chromatogram (EIC) for (a) laboratory generated sample from α-pinene/O_3 and (b) pooled night-time ambient summer sample from Hyytiälä, Finland collected during 16 to 25 August 2011. The chromatographic peaks correspond to (1) diaterpenylic acid with *m/z* 171.0662, (2) *cis*-pinic acid with *m/z* 185.0818, (3) *cis*-caric acid with m/z 185.0818, (d$_1$) dimer with m/z 377.1454, (d$_2$) dimer with m/z 357.1552, (d$_3$) dimer with *m/z* 367.1762 and (d$_4$) dimer with *m/z* 337.1652. In the ambient sample (b) only one dimer, d$_2$, was observed; the other small peaks between 36 and 42 min retention time in (b) do not correspond to d$_1$, d$_3$ and d$_4$.

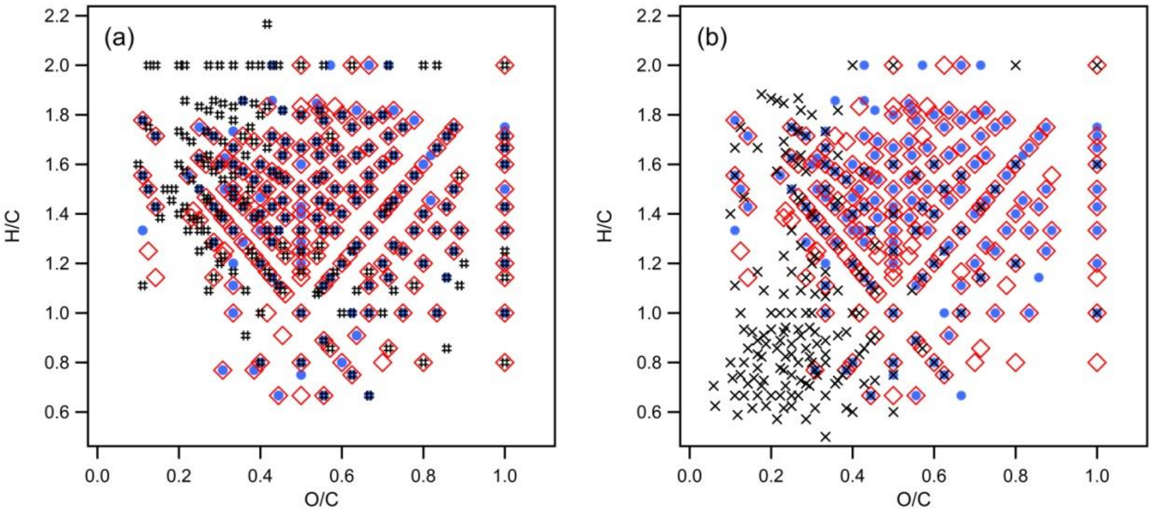

Figure 3. Van Krevelen diagrams for compounds containing only carbon, hydrogen and oxygen in the samples from (a) α-pinene/O_3, BVOC mixture/O_3 and the boreal forest, Hyytiälä, Finland; and (b) *α*-pinene/O_3, BVOC mixture/O_3 and Tivoli Industrial Estate and Docks, Cork, Ireland. Filled blue circles correspond to species from α-pinene/O_3 experiment, red diamonds to BVOC mixture/O_3 experiments, black hashes to the Hyytiälä samples and black crosses to the Tivoli Industrial Estate and Docks, Cork samples.

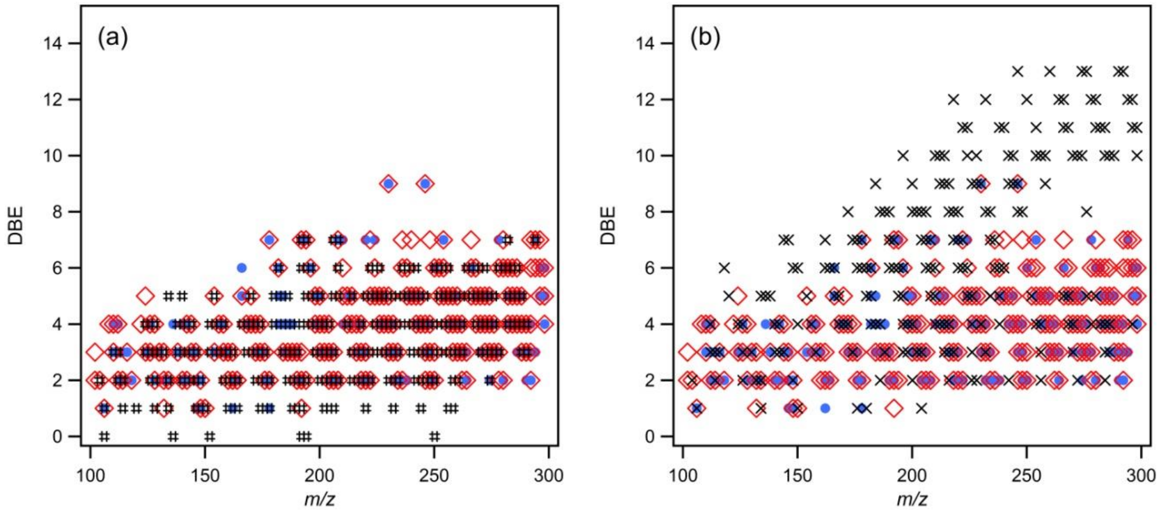

Figure 4. Double bond equivalents (DBE) vs. mass to charge (*m/z*) ratios for all compounds containing only carbon, hydrogen and oxygen in the samples from (a) *α*-pinene/O_3, BVOC mixture/O_3 experiments and boreal forest, Hyytiälä, Finland, and (b) *α*-pinene/O_3, BVOC mixture/O_3 experiments and Tivoli Industrial Estate and Docks, Cork, Ireland. Filled blue circles correspond to species from α-pinene/O_3 experiment, red diamonds to BVOC mixture/O_3 experiments, black hashes to the Hyytiälä samples and black crosses to the Tivoli Industrial Estate and Docks, Cork samples.

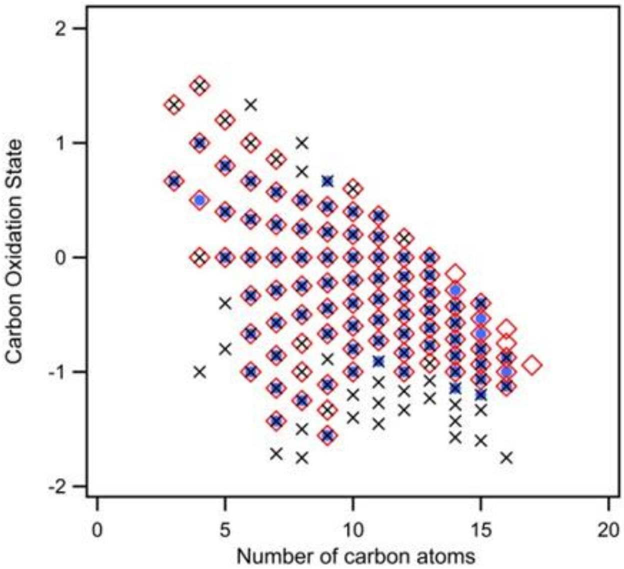

Figure 5. Carbon oxidation state for molecules containing only carbon, hydrogen and oxygen in the *α*-pinene/O_3 experiments (blue circles), the BVOC mixture/O_3 experiments (red diamonds) and the boreal forest, Hyytiälä, Finland (black crosses).

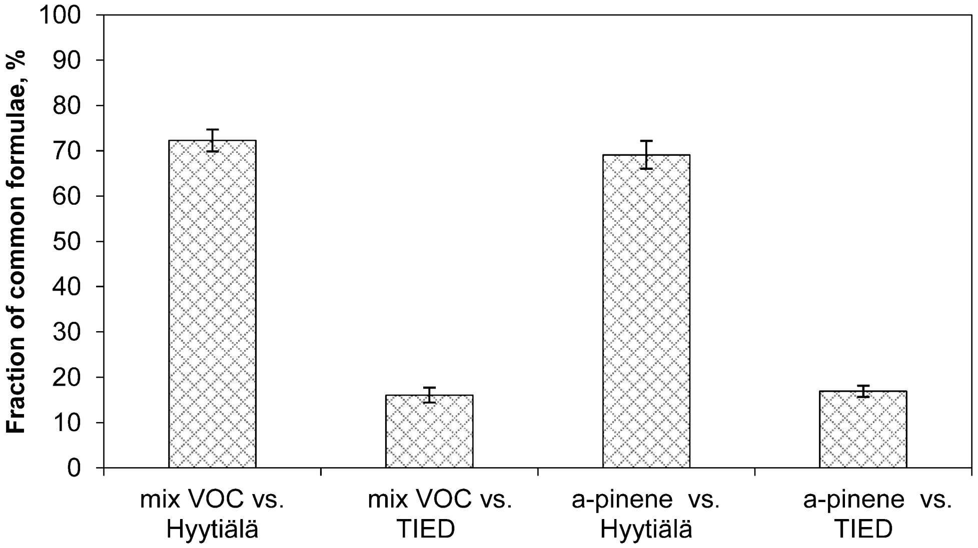

Figure 6. Fraction of molecular formulae below 300 Da found in SOA from both laboratory-generated and ambient samples relative to the total number of formulae in the ambient samples. The error bars represent the standard deviation of three replicate chamber experiments.

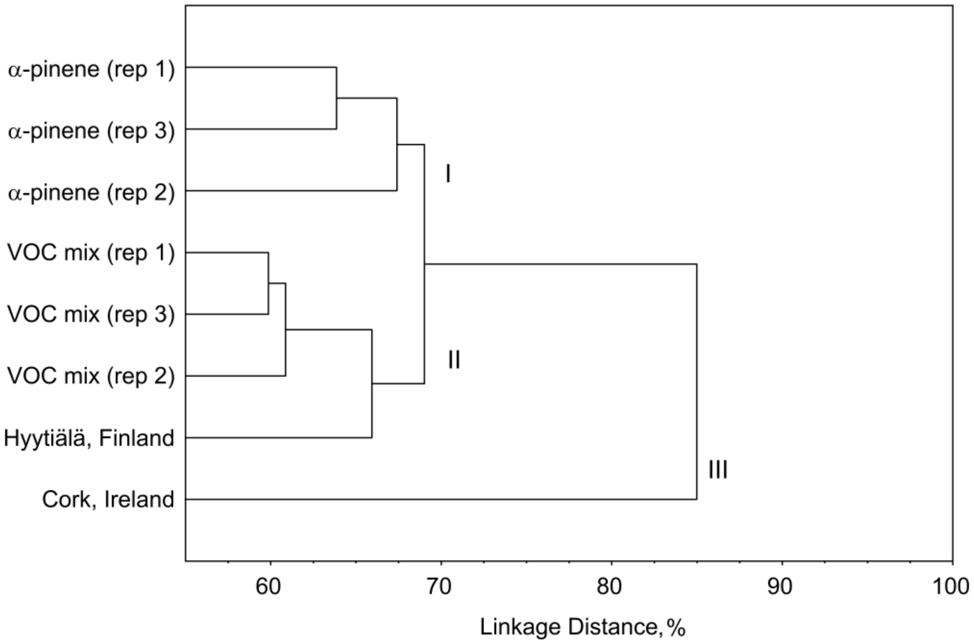

Figure 7. Tree diagram derived from hierarchical cluster analysis (HCA) of ambient and laboratory-generated samples (unweighted pair-group average linkage method). The linkage distance between two samples, expressed in percentage, has been calculated as the difference between the total ions considered in the cluster analysis (451 ions, 100%) and the number of common ions between the samples (see text for detailed explanation). Rep 1, 2 and 3 correspond to chamber replicate experiments.