

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

**by**

K. P. Wyche *et al.*

The authors are grateful to both referees for their excellent, thoughtful and insightful reviews, their comments were most welcome and indeed were very useful for manuscript improvements to be implemented. The following document addresses all of the points raised by both reviewers and explicitly details all of the correction made to the manuscript (revised manuscript attached separately with “track changes” documenting alterations) in accordance with both of the referee’s suggestions.

**Author response key:**

Plain, green text denotes referee comments

*Italicised, back text, prefixed with “Author Response:” denotes direct author response to preceding referee comment*

Plain, blue text, inside " " marks, denotes author changes to the manuscript in response to requests or recommendations from the reviewer

**Anonymous Referee #1**

Received and published: 26 February 2015

The manuscript of Wyche et al. represents an exciting step forward in the analysis of complex data sets, relevant to the formation of secondary organic aerosol (SOA). The authors use three statistical analysis approaches (PCA, HCA, and PLS-DA) for dimension reduction. The methods are applied to data obtained from a number of chamber studies involving different precursors grouped into four categories; application of the methods reproducibly resulted in successful classification of the gas- and particle-phase composition spectra by precursor category. While the authors suggested the potential for such approaches in ambient data mining, the results were not overstated and potential limitations were acknowledged. The paper was well written and the methodology was easy to follow. This paper should be of great interest to ACP readers involved in acquisition and interpretation of such data sets, as well as to readers interested in improving model representation of SOA.

*Author Response: The authors thank the referee for their summary and recommendation.*

---

Content

Studies have shown that limonene for example typically has higher SOA yields than other cyclic monoterpenes (Lee et al., 2006, JGR 111 D17305; Fry et al., 2014, ES&T 48: 11944-11953). Though the oxidation products of the two cyclic monoterpenes were separable from the other categories by the statistical approaches, did the results also suggest differences between the two precursors that may help explain such observations?

*Author Response: To a certain degree the monoterpenes, limonene and  $\alpha$ -pinene were separable within the monoterpene group, with three out of the four  $\alpha$ -pinene experiments located to the upper and right region of the cluster. The referee is correct, this distribution within the statistical space results from the differences in respective gas-phase organic compositions of the two precursor types, which then in turn impact SOA yield and composition. As such the text has been amended in*

Section 5, such that there are now 3 distinct discussion sections: 5.1 Mapping chemistry (the original text), 5.2 Mapping within a class (i.e. the following text supplied in response to this specific comment by referee 1) and 5.3 Mapping reactivity (i.e. additional text supplied in response to the following comment from referee 1 – see below – and comment 14) from referee 2). The additional text has been inserted on page 1676, after line 2, to answer the referees' question and to include the references that the referee has suggested:

*"5.2 Mapping within a class*

Within the monoterpene group there is a small degree of separation between the limonene and  $\alpha$ -pinene experiments, with three out of the four  $\alpha$ -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,  $\alpha$ -pinene and limonene react at somewhat different rates with respect to both OH and O<sub>3</sub> (Atkinson and Arey, 2003). 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  $\alpha$ -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"

---

Along the same lines, within a category, does the spread/distance of clusters/specificity indicate dependence on VOC/NO levels, RH, or other factors thought to influence SOA formation?

*Author Response:* This is another interesting point, which we were originally intending to reserve for future work. However, following this comment and comment

13) from referee 2, we have now included a small separate analysis of the toluene data, which was conducted over a reasonable range of VOC/NO<sub>x</sub> conditions (low, medium, high NO<sub>x</sub>), to demonstrate how the experiments of one specific precursor type distribute/separate within the original class on account of system driving specifics, in this case, VOC/NO<sub>x</sub> ratio. This new information has been included in a new section, Section 5.3, following the original text on page 1676, after line 2, along with an additional Figure (10), shown here:

### “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<sub>x</sub> ratios. The VOC/NO<sub>x</sub> ratios employed nominally represent “low”, “medium” and “high” NO<sub>x</sub> conditions, with values of roughly 11 (i.e. low NO<sub>x</sub>, NO<sub>x</sub> limited; two experiments), 4 (i.e. moderate NO<sub>x</sub>, two experiments) and 1 (i.e. high NO<sub>x</sub>, VOC limited; one experiment), respectively (e.g. see work by Wagner et al., 2003). 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 oxidation spectra distribute in statistical space according to their respective VOC/NO<sub>x</sub> ratios. Figure 10 shows the low NO<sub>x</sub>, high VOC/NO<sub>x</sub> 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<sub>x</sub> levels in the NO<sub>x</sub> limited regime, result in low [O<sub>3</sub>] (via NO<sub>2</sub> → O(<sup>3</sup>P) photolysis) and low [OH] (via O<sub>3</sub> → O(<sup>1</sup>D) photolysis). Similarly, the relatively large contribution from cresol to the low NO<sub>x</sub> 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  $\text{NO}_x$ , medium  $\text{VOC}/\text{NO}_x$  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 than that formed under low  $\text{NO}_x$  conditions. Larger net benzaldehyde concentrations originate from greater system reactivity and greater abundance of  $\text{NO}$  to fuel the  $\text{RO}_2 + \text{NO}$  reaction.

The high  $\text{NO}_x$ , low  $\text{VOC}/\text{NO}_x$  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  $\text{NO}_x$  experiments) and the resultant presence of ring-opening product ions (as the case for the moderate  $\text{NO}_x$  experiments). However, the low  $\text{VOC}/\text{NO}_x$  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  $\text{NO}_x$  conditions, owing to reaction through the  $\text{RO}_2 + \text{NO}$  channel)."

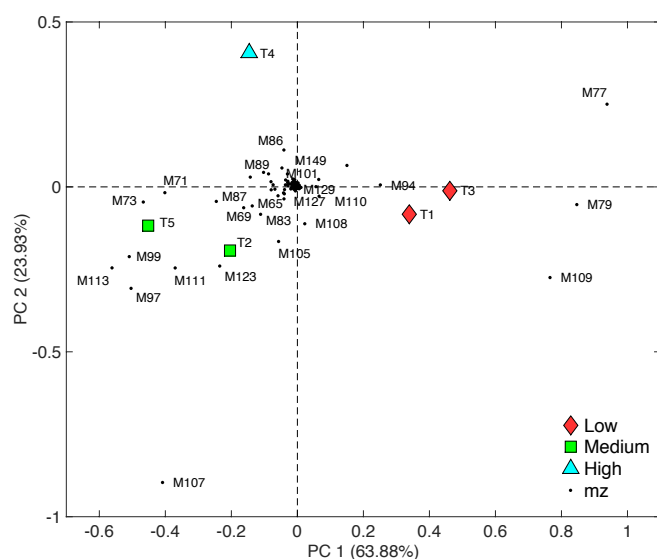


Figure 10: PCA loadings bi-plot of the second vs. first principal components derived from the PCA analysis of the toluene experiments. Experiments were conducted under low NO<sub>x</sub>, high VOC/NO<sub>x</sub> ratio (red diamonds), moderate NO<sub>x</sub>, medium VOC/NO<sub>x</sub> ratio (green squares) and high NO<sub>x</sub>, low VOC/NO<sub>x</sub> 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.

---

Lee et al. reported a number of m/z fragments for oxidation products of isoprene, monoterpenes, and a sesquiterpene detected by PTR-MS. It would be interesting to know if the same conclusions could be ascertained from the reported m/z fragments in Lee et al.; namely that there are significant relationships between the fragments of “like” precursors. There may not be enough data to perform a full analysis as was done in the subject manuscript, but the authors may think about whether their results are similar/supported by Lee et al.

*Author Response:* The referee is indeed correct, there are similarities between the fragments of products of “like” precursors. Here, the main similarities are between the monoterpenes limonene and  $\alpha$ -pinene, and myrcene and linalool. However, we have highlighted this information for the major oxidation products in Section 4.2, on page 1666, line 15 onwards, e.g. “The monoterpene groupings are influenced by the presence of m/z 107, 151 and 169 (primary aldehydes, pionaldehyde and limononaldehyde) and 139 (primary ketone, limonaketone) ions in their mass spectra” and “Helping to separate the straight chain from cyclic monoterpenes are m/z 95 and 93, 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)”

However, following this comment from referee 1 and Technical Comment 5) from referee 2, the text has been amended as follows on page 1666, line 28, to link the findings of this work to that of Lee et al., 2010:

“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.”

---

It is suggested that the authors carefully review their citations. There are several places in which widely cited previous work is omitted in favor of self-citations.

*Author Response:* We thank the referee for pointing this out and we have amended our citations as follows, however we would like to highlight that in some instances we include numerous citations to our previous work in order to explain specifics of the experiments that are employed here.

---

One example can be found on p. 1655, line 22, regarding the importance of findings from chamber studies over the past decade. The authors do not cite the work coming out of the Caltech chamber (as early as the late 1990's); the parameters from Hoffman et al., 1997 and Griffin et al., 1999 still are widely used in chemical transport models (see also papers by Cocker, D.).

Restricted to the last 10 years, lead authors on Caltech chamber publications include: Ng, N.; Kroll, J.; Surratt, J. and others.

Much influential work has also been conducted over the last 10 years in the Carnegie Mellon chamber; lead authors include: Presto, A., Hildebrandt, L., Grieshop, A., and others.

*Author Response:* The referee is indeed correct, this is an oversight on our part and the following citations and references have been added to the amended text on,

*Page 1655, line 22:*

“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)."

*Page 1658, line 6:*

"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)."

*Page 1665, line 6:*

"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)."

*Page 1670, line 9:*

"For further details regarding  $\beta$ -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."

*Page 1670, line 18:*

"For further details regarding  $\alpha$ -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.”

*Page 1675, line 12:*

“ $\beta$ -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)”

*References added to manuscript list:*

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.

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.

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.

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.

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.

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.

Jaoui, M. and Kamens, R. M.: Gaseous and particulate oxidation products analysis of a mixture of  $\alpha$ -pinene +  $\beta$ -pinene/O<sub>3</sub>/air in the absence of light and  $\alpha$ -pinene +  $\beta$ -pinene/NO<sub>x</sub>/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.

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 PM<sub>2.5</sub>, *J. Atmos. Chem.*, 47, 79–100, 2004.

Presto, A. A., HuffHartz, K. E., and Donahue, N. M.: Secondary Organic Aerosol Production from Terpene Ozonolysis. 2. Effect of NO<sub>x</sub> Concentration, *Environ. Sci. Technol.*, 39, 7046–7054, 2005.

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.

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<sub>x</sub> and their Detection in Ambient PM<sub>2.5</sub>, *Environ. Sci. Technol.*, 40(12), 3819–3828, doi: 10.1021/es052566z, 2006.

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.

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/NO<sub>x</sub> photooxidation, *Geophys. Res. Lett.*, 33, L13805, doi:10.1029/2006gl026523, 2006.

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.

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.

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.

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.

Winterhalter, R., Herrmann, F., Kanawati, B., Nguyen, T. L., Peeters, J., Vereecken, L., and Moortgat, G. K.: The gas-phase ozonolysis of β-caryophyllene (C<sub>15</sub>H<sub>24</sub>). Part I: an experimental study, *Phys. Chem. Chem. Phys.*, 11, 4152–4172, 2009.

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  $\beta$ -caryophyllene. *Atmos. Chem. Phys.*, 11, 1735–1751, doi:10.5194/acp-11-1735-2011, 2011.

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.

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.

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  $\beta$ -caryophyllene ozonolysis, *Atmos. Chem. Phys.*, 11, 121–132, doi:10.5194/acp-11-121-2011, 2011.

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

---

#### Editorial

p 1655, line 18: “Atmospheric chemistry” as a system is awkward. One suggestion: “The chemistry of the atmospheric system is highly nonlinear. . .”

*Author Response: The text has been altered accordingly*

---

p 1656, line 9: The “CIR-TOF-MS” abbreviation can be introduced on line 4 and used exclusively here.

*Author Response: The text has been altered accordingly*

---

p. 1658, section 2 heading: The adjective “Experimental “ needs a verb, or needs to

350 be changed to a noun.

351 *Author Response: The text has been altered to read "Experimental Details"*

---

352

353 p. 1661, line 8: Remove "/" prior to photolytic

354 *Author Response: The text has been altered accordingly*

---

355

356 p. 1666, line 19: Change "caryophyllon" to "caryophyllene"

357 *Author Response: The text has been altered accordingly*

---

358

359 p. 1667, line 26: It is not clear what is meant by "oxidized atmospheres".

360 *Author Response: For clarity, the text has been altered to read,*

361 "As a further test of the technique to distinguish between and to classify VOCs and  
362 the matrix of oxidized organic compounds that may derive from their atmospheric  
363 chemistry, test data from an anthropogenic system was introduced into the model."

---

364

365 p. 1669, line 10-13: It is suggested the sentence be reworded to more clearly  
366 indicate that data are available for only two experiments per each of the two  
367 noncyclic monoterpenes. As written the focus is on "only two types of precursor",  
368 which is the same as the cyclic monoterpenes. In summary, it is the number of  
369 experiments, not the number of precursors that is likely affecting the result.

370 *Author Response: the text has been altered to read,*

371 "The greater spread in confidence of the noncyclic monoterpene group is once again  
372 likely to result to some extent from the low number of repeat experiments  
373 employed (i.e. only two each for myrcene and linalool)."

---

374

375 Can the panel and text sizes be increased for Figs. 7 and 8? They are hard to read  
376 (even w/significant expansion-175%)?

377 *Author Response: Figures 7a, 7b, 8a and 8b have been altered as requested, with*  
378 *increased panel and text sizes. We will also work with the editing team to ensure in*  
379 *typesetting that all figures are sufficiently clear.*

---

380

381 The authors may consider more clearly indicating the difference between Figs. 7 and  
382 8 in the figure panels (e.g., just adding a AMS and LC-MS header).

383 *Author Response: Figures 7a, 7b, 8a and 8b have been altered as requested, with*  
384 *additional headers to indicate LC-MS/MS or cTOF-AMS data.*

---

385

386 The authors may consider using consistent colors for the different categories in each  
387 of the plots (generally the same in PCA plots, but not in dendorgrams).

388 *Author Response: We thanks the referee for pointing this out, but we would prefer to*  
389 *leave the dendrogram colors as there are, as they represent the best combination for*  
390 *the clearest description of the data, i.e. the dendrogram colors were chosen*  
391 *specifically to optimize clarity of view within the figure. The colors and symbols are*  
392 *the same for all PCA and PLS-DA plots, e.g. isoprene is always represented by a blue*  
393 *diamond.*

---

394

395 Caption Figure 10: Remove "/" after abundance

396 *Author Response: The text has been altered accordingly*

---

397

**Anonymous Referee #2**

Received and published: 10 March 2015

Overview:

Wyche et al. describe development of a chemometric mapping of gas-phase and particle-phase matrices from oxidation of several relevant compounds: isoprene, terpenes (i.e. linalool, myrcene, limonene,  $\alpha$ -pinene, B-caryophyllene), toluene, as well as the oxidation products of fig and birch trees. This is done via principal component analysis (PCA), hierarchical cluster analysis (HCA), and positive least squares-discriminant analysis (PLS-DA) on the mass spectra of the oxidation products from several systems. PCA is performed on gas-phase observations of chamber data using chemical ionization reaction time-of-flight mass spectrometry for the gas-phase to separate distinct regions associated with oxidation of isoprene, cyclic monoterpenes, B-caryophyllene, single chain monoterpenes, and toluene. Addition of mesocosm data (plant emission oxidation experiments) shows expected mapping of oxidation products from fig and birch trees as related to the expected emissions being isoprene-dominated and cyclic monoterpene-dominated, respectively. HCA analysis is performed, supporting the relation of the mesocosm oxidation products with associated precursor oxidation schemes represented in chamber data. Similar PCA and HCA analysis applied to the particle-phase data (liquid chromatography-ion trap mass spectrometry and compact time-of-flight aerosol mass spectrometry) also shows separation by precursor type. Using a zero-dimensional box model simulation based on the Master Chemical Mechanism for  $\alpha$ -pinene photooxidation, it is argued that the employed statistical deconvolution techniques could be applicable for determining precursor type and potential mechanisms from ambient data on the basis that “model mass spectra” simulated under increasingly complex (closer to ambient conditions) are generally well-captured by “model mass spectra” under conditions more like a typical chamber experiment.

Reviewer’s recommendation:

The article overall is well-written and is novel in that it seems to provide great

potential for use as identifying chemistry from a particular precursor type given complex sets of mass spectra from oxidation products in the gas and particle phases. The weakest argument is that made for use of this technique to elucidate chemical mechanisms in addition to identifying precursor type, as the specificity for separating mechanistic pathways is not fully demonstrated in the current analyses. Still, the content is appropriate for ACP readers and I would recommend publication after the following comments are addressed.

*Author Response: The authors thank the referee for their summary and recommendation.*

---

General Comments:

1) Abstract: The reader would benefit from more specific result statements included here, similar to the lines included in the Introduction section p. 1657, lines 7-15.

*Author Response: The following text has been added as requested,*

*Abstract, line 18, page 1653,*

*"Results show that "model" biogenic oxidative systems can be successfully separated and classified according to their gaseous oxidation products."*

*Abstract, line 26, page 1653,*

*"More specifically, the addition 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 biogenic sources and controlling oxidative chemistry."*

---

2) P. 1656, lines 20-25: Discussion of currently used statistical techniques is rather cursory. There is not even mention of the commonly used positive matrix factorization technique widely used for AMS data analysis. Additional details of why the presented technique is novel/necessary should be discussed.

*Author Response: We have amended the introduction as follows; to include a more*

*detailed discussion of other more commonly used statistical techniques, including positive matrix factorization. The additional details regarding why the technique presented here is novel are addressed in reply to Referee 2, point 3) below:*

“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



487 analysis of each of the sets of measurements on their own.

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

---

496

497 3) The authors mention p. 1657, lines 16-25, the potential for this technique to be  
498 used on ambient data sets, but the paper would be stronger references were  
499 provided for which similar statistical analyses are being done to map out oxidation  
500 chemistry related to certain precursors as is done here, or utilize available and  
501 published field data in the described analyses of the paper to prove the point. For  
502 example, why would this technique be useful over positive matrix factorization  
503 techniques, which now includes a way of identifying SOA formation contributions  
504 from isoprene?

505 *Author Response: The referee is indeed correct; the paper would be stronger if we*  
506 *could have included ambient data in the statistical analyses in the same way that we*  
507 *included the mesocosm data. Unfortunately, we do not have such data available and*  
508 *have tried to address the issue to some extent using the MCM simulations to model a*  
509 *representatively complex system. It should be noted that we could find no useable*  
510 *published field data (as the referee suggests). A large amount of such data*  
511 *comprises solely the compounds of interest and not the detailed mass spectra, and*  
512 *where mass spectra were available, they were incompatible/not fit for purpose,*  
513 *either being e.g. flux data and/or comprising PTR-Quadrupole-MS data recorded in*  
514 *selected ion mode.*

515 *We feel (as was pointed out by referee 1) that we have been fully open regarding the*  
516 *potential limitations of this methodology and have attempted to address them to the*

517 *best of ability.*

518 *We feel that taking the technique from first principle, to single precursor and then to*  
519 *mesocosm scenarios, achieves the goals that were set out and that application to the*  
520 *“real world” would represent the next logical step, for our research and for other*  
521 *colleagues; given the results presented in this proof of concept study, using this*  
522 *approach successfully in the field is the ultimate goal.*

523 *Regarding Positive matrix factorization (PMF) and the uniqueness of the method*  
524 *presented here; PMF is an extremely valuable technique now used successfully for*  
525 *PM source apportionment. The ensemble methodology described here however is*  
526 *not intended as a replacement, far from it, it simply presents a different approach to*  
527 *isolate composition (and potentially therefore, sources) and moreover to identify and*  
528 *map different chemical pathways. PMF generally utilizes data solely from particle*  
529 *phase measurements (however, PMF with both gas and particle phase data have*  
530 *been reported, i.e. Slowik et al., ACP, 2010) and hence usually from one analytical*  
531 *technique. However, the ensemble method discussed here utilizes data from both the*  
532 *gas- and particle-phases and data from different analytical techniques. Furthermore,*  
533 *it is not designed simply to comprise a type of source-receptor/source apportionment*  
534 *model; the objective is rather compositional isolation to instruct on potential sources*  
535 *AND on underlying chemical processes.*

536 *The methodologies described here are fundamentally simpler to use and require*  
537 *fewer base assumptions (there is less importance placed on user skill to determine*  
538 *the correct input parameters, number of factors to employ etc); the techniques work*  
539 *on fewer data, require no reference datasets and no prior knowledge of sources is*  
540 *required.*

541 *We thank the referee for this useful comment and have altered the introduction to*  
542 *highlight these distinctions, with the addition of the following text to line 6, page*  
543 *1657:*

544 *“Unlike other statistical techniques such as PMF, the ensemble methodology*  
545 *presented here does not require the use of additional external databases*  
546 *(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.”

---

4) P. 1659, lines 9-12: While the precursors were reacted to near completion in the chamber experiments, how can one verify that the oxidation scheme went to completion to match, for example, major oxidation products in the MCM schemes presented in Figure 9. That is, should not theoretically the PCA analysis for the chamber experiments ideally match the MCM PCA if the chamber experiments are covering the same range as the simulated chemistry?

*First of all, we must reiterate as was pointed out in the manuscript, the MCM simulations are not idealized, and their inclusion is not for the purpose of making a direct model-measurement comparison. Owing to a lack of in field ambient spectra, the simulations are designed to exemplify how the gas phase spectra could change if the system complexity was increased. The model is not included for direct comparison with the measurements, it is intended as a guide to help bridge the gap between what we have reported from “model” and “mesocosm” experiments and the real world. This is stated on page 1677, line 16.*

*However, we can verify that precursor oxidation went to completion to match the measurements, because the model was purposefully constrained to measured  $\alpha$ -pinene concentration, i.e. the primary reactivity in the model was constrained to match the primary reactivity in the measurements. Furthermore, from our previous work (see Camredon et al., 2010) the modeled and measured temporal profiles of the precursor and first (and later) generation products are in reasonably good agreement, suggesting that the reactivities of the systems are comparable (i.e. that the model system is chemically dynamic in a similar way as you would see in the chamber). Use of the MCM under such single precursor chamber conditions has been evaluated in details on several previous occasions, for example see work by Bloss et al., 2005 (toluene), Camredon et al., 2010 – supplementary information ( $\alpha$ -pinene), Rickard et al., 2010 (1,3,5-TMB) and Jenkin et al., 2012 ( $\beta$ -caryophyllene). The reader*

578 has now been informed of this with the addition of the following text to page 1677,  
579 line 18:

580 “For work regarding the evaluation of the MCM with respect to single VOC precursor  
581 chamber experiments (including model-measurement intercomparison), see for  
582 example, Bloss et al., 2005 (toluene), Metzger et al., 2008 and Rickard et al., 2010  
583 (1,3,5-TMB), Camredon et al., 2010 ( $\alpha$ -pinene) and Jenkin et al., 2012 ( $\beta$ -  
584 caryophyllene).”

---

585

586 5) P. 1659, Section 2.2: What impacts on the comparison of the chemical oxidation  
587 would be expected for the systems, considering that these systems were run across  
588 several environmental chambers?

589 *Author Response: Each (Teflon) chamber system will have its own specific*  
590 *background wall chemistry, which arguably changes by various degrees from day-to-*  
591 *day and even between experiments (i.e. cleanliness of chamber will vary). For certain*  
592 *chambers, e.g. EUPHORE (Bloss et al., 2005; Zador et al., J. Atmos. Chem., 55:147-*  
593 *166, 2006) and PSISC (Metzger et al., 2008), detailed auxiliary mechanisms have been*  
594 *constructed and tested. These chemical mechanisms are based on similar reactions*  
595 *and fundamental properties and go some way towards explaining the intricacies of*  
596 *environment chamber experiments. From our experience, where rates of reaction*  
597 *(influenced by radical chemistry associated with the chamber walls) can change*  
598 *between chambers for a give set of experimental conditions, the overall composition*  
599 *of the gas phase does not vary widely for the types of experiments studied. This is*  
600 *inherent in the data reported here, as the spectra collected are compositionally*  
601 *similar between chambers. The primary additives generated from the chamber walls*  
602 *that could influence organic gas phase composition are nitrogen-containing species*  
603 *such as HONO. However, for the experiments reported here (with the exception of*  
604 *the mesocosm experiments), initial  $[NO_x]$  added >  $[NO_x]$  liberated from chamber*  
605 *walls.*

606 *In addition to the wall chemistry, the light intensity will vary from chamber to*  
607 *chamber, however, EUPHORE employs natural solar radiation and the MAC and PSISC*

are designed to approximate the solar spectrum, as detailed in the work referenced appropriately in the experiment section 2.2, page 1659.

In summary, we expect the impacts to be fairly minimal and within acknowledged uncertainty bounds.

---

6) P. 1667, lines 2-4: Aromatic compounds are also observed to be emitted from the biosphere, (e.g. aromatics like toluene as well as aromatic monoterpenes; see Guenther et al., 2012 and references within; doi:10.5194/gmd-5-1471-2012). Would the authors expect addition of these compounds to map onto the same region as toluene/monoterpenes, and would this not change the interpretation that the analysis can separate anthropogenic and biogenic precursor sources vs. just precursor structure? Do the authors have potential ideas for why there is overlap with toluene and the single chain monoterpenes? Would single chain sesquiterpenes (e.g. farnasene) also map out separately from B-caryophyllene as the cyclic and straight chain monoterpenes do?

**Author Response:** Guenther et al., (2012) estimated global annual emissions of  $\sim 535 \text{ Tg yr}^{-1}$  for isoprene and a sum of  $\sim 147 \text{ Tg yr}^{-1}$  for non-aromatic terpenes; total sesquiterpene emissions were of the order  $29 \text{ Tg yr}^{-1}$ . In contrast aromatic terpene emissions were estimated in a group of  $\sim$  thirty “other” monoterpenes with total annual emissions  $\sim 15 \text{ Tg yr}^{-1}$ , i.e. only 2 % of  $\Sigma(\text{isoprene, major non-aromatic terpenes and sesquiterpenes})$ . Estimated annual toluene emissions are not individually reported by Guenther et al., but are assumed to be relatively minor, reported with an “additional 11 stress VOC(s)”, with a total emissions value of  $7.8 \text{ Tg yr}^{-1}$ . Sindalevora et al. (2014) report mean isoprene,  $\Sigma(\text{monoterpenes})$ ,  $\Sigma(\text{sesquiterpenes})$  and toluene annual global emissions to be of the order 594, 95, 20 and  $1.5 \text{ Tg yr}^{-1}$ , respectively, i.e. toluene emissions are only  $\sim 0.2 \%$  of  $\Sigma(\text{isoprene, major non-aromatic terpenes and sesquiterpenes})$ . In summary, biogenic aromatic compounds comprise only a small fraction of total average terpene emissions and it is likely therefore that their oxidation products would be present in such low abundance compared to those discussed within our work, that they would not cause

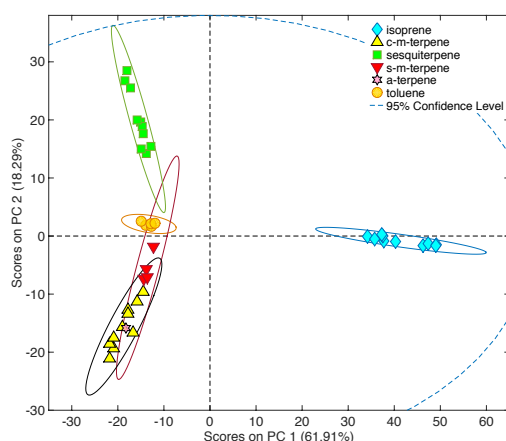
638 severe spectral/statistical interference.

639 Under exceptional circumstances where aromatic terpenes are present in significant  
640 abundance, it is possible that their compositional spectra could group in the vicinity  
641 of those of other monoterpenes, on account of them possessing spectra with some  
642 similar features, e.g. Lee et al. (2006; DOI: 10.1029/2006JD007050) investigated the  
643 PTR-MS spectra of methyl chavicol and reported the presence of a parent ion of  $m/z$   
644 149 and major product ions of  $m/z$  151, 137, 121 and 109. Of these ions  $m/z$  149, 151  
645 and 137 do not appear in toluene PTR-MS spectra and moreover, the ions of  $m/z$  151,  
646 137 and 109 are also significant contributors to monoterpene PTR-MS oxidation  
647 spectra.

648 Experiments conducted with aromatic terpenes did not form the focus of this initial  
649 work and as such no data is available here to test the exact positioning of aromatic  
650 terpenes in the statistical space, however, in addition to the findings of Guenther,  
651 Sindalevora and Lee et al., the hypothesis that aromatic “like” BVOCs would not  
652 group with toluene was tested here with data from  $\alpha$ -terpinene photooxidation.  $\alpha$ -  
653 terpinene is not an aromatic terpene, but unlike  $\alpha$ -pinene and limonene it does  
654 contain a 6-member carbon ring containing 2 C=C double bonds.

655 Results:  $\alpha$ -terpinene oxidation produces a significant amount of  $\alpha$ -terpinaldehyde,  
656 which like pinonaldehyde (from  $\alpha$ -pinene photooxidation) possesses a spectral  
657 fingerprint including ions of  $m/z$  169, 151, 123 and 107. As with other non-aromatic  
658 terpenes,  $\alpha$ -terpinene photooxidation PTR-MS spectra also contains ions of  $m/z$  139  
659 (c.f. limonaketone from limonene) and 171 (also observed in  $\alpha$ -pinene photooxidation  
660 spectra), as well as more unique features, such as an  $m/z$  143 ion of significant  
661 abundance (also see Lee et al., 2006, DOI: 10.1029/2006JD007050). In combination,  
662 such features help to group the  $\alpha$ -terpinene photooxidation spectrum in the vicinity  
663 of other terpenes, separate in statistical space from the aromatic toluene  
664 photooxidation spectra.

665



666

667

668 *The above figure and a short description matching the above text have been added*  
 669 *to a new Supplement document and section 4.3 text has been amended accordingly*  
 670 *in order to direct the reader appropriately:*

671 *“A brief discussion regarding aromatic BVOCs is provided separately in the*  
 672 *supplementary material.”*

673

674 *We do not believe that there is significant overlap between the straight chain*  
 675 *terpenes and toluene; the overlap observed is from the high (95%) confidence limits,*  
 676 *the magnitude of which is discussed in the main text of the paper. Also as discussed*  
 677 *in the paper, the toluene cluster is located close to the straight chain monoterpene*  
 678 *cluster owing to similar (but not identical) ion content (but not abundance), including*  
 679 *primarily  $m/z$  93. Toluene separates from the straight chain monoterpenes owing to*  
 680 *unique spectral features such as  $m/z$  99 and 113, resulting from ring-cleavage and*  
 681 *the subsequent production of e.g. angelica lactone and hexenedione (and isomers*  
 682 *thereof), respectively.*

683 *We have not yet performed experiments with single chain sesquiterpenes such as*  
 684 *farnasene, however it is likely that, as with the cyclic and non-cyclic monoterpenes,*  
 685 *single chain sesquiterpene spectra would group in the vicinity of cyclic*

686 sesquiterpenes, owing to their similar composition and chemistry, but uniquely  
687 displaced as a consequence of their structure (e.g. unsaturated, double bond  
688 content).

---

689

690 7) P. 1669, lines 17-20: Why are the birch trees sensitivity cited to be low due to only  
691 two repeat experiments, but the fig trees experiments also only having two repeat  
692 experiments have high sensitivity and specificity?

693 *Author Response:* As stated in the main text, page 1669, Section 4.4, lines 17 – 20,  
694 the low birch sensitivity may result from a combination of (i) a low number of repeat  
695 experiments AND (ii) the aging trees producing lower emissions during the last  
696 experiment. As we state in the current text, the low sensitivity for birch trees is likely  
697 to result from a combination of these two factors.

---

698

699 8) Section 4.5: Why do the authors choose to do separate PCA analyses on the gas-  
700 phase data versus the particle-phase data? Considering that there would be overlap  
701 for semi-volatiles (e.g. B-caryophyllinc acid as stated p. 1670, line 10), should not a  
702 PCA analysis (though mindful of the different analysis technique) be more ideal (as  
703 well as for interpreting ambient data) to map regions of any chemistry associated  
704 with various precursor types? Should it be inferred that the separation of analysis on  
705 gas vs. particle-phase data is the recommended approach to apply this type of  
706 analysis to ambient data? I would be curious to see how the analyses would compare  
707 between gas-phase only, particle-phase only, and gas and particle-phase together.

708 *Author Response:* What the referee suggests here is an excellent idea, and one that  
709 we were hoping to execute in future work, however, such an undertaking will be  
710 extremely complex and a step beyond what we are trying to present here. To begin  
711 with, this task would require knowledge of (i) ALL of the species present in both the  
712 gas and aerosol phases (which the community does not possess yet) and (ii) the  
713 EXACT fragmentation patterns and ionisation routes for all of these species under  
714 each of the ionisation regimes used. If these data were available, one would then  
715 have to (re-) construct mass spectra representing the gas and particle phase matrices



for each of the experiments employed, with no detection method specific alteration to the mass of each parent compound (e.g. protonation, adduct formation, fragmentation etc) or instrument biases. At present it is not possible to do this accurately, however we are steadily working towards obtaining such knowledge. Moreover, methodology such as Parallel Factor Analysis would perhaps be more suitable to perform such an investigation, which again is a whole new piece of work in itself. The purpose of this work is to introduce these new methods of working with large atmospheric data sets, partially the aim of which is to reduce computational load, not extend it, which would be required in this instance.

---

9) Figure 7a: The plot is produced using LC-MS particle data, but what biases might be at work here based on this technique's sensitivity to certain organic aerosol products?

**Author Response:** In any analytical method there will be a degree of bias towards particular species. The LC-MS was operated in negative ionisation mode and so here we are looking at the species with acidic protons that can be removed easily. This leads to certain biases, as the ionisation efficiency is different for different structures. However, this plot shows the key ions that allow the mass spectral patterns of the SOA from different precursors to be separated based on the observed chemical speciation. The same analysis in positive mode (which targets different functional groups) will likely lead to some different ions being important for differentiating the precursors. However all samples were analysed using the same methods and so the impact of any systematic biases are minimised.

---

10) Figure 8a and accompanying discussion p. 1671, lines 15- p. 1672, line 3: The meaning of M43, M44 should be discussed in context as to whether the relative aging of the oxidation systems were actually comparable. Although the precursors were near complete reaction, did the oxidation develop enough such that B-caryophyllene schemes could take on additional spectral contributions from M43, M44? That is, was their sufficient oxidant such that secondary reactions (producing

746 characteristic oxidation products and associated ions) could occur equally across  
747 precursor systems?

748 *Author response:* the meaning of M43 and M44 has been discussed in the text (e.g.  
749 see page 1671, line 15 onwards). As can be seen from the Experimental section (and  
750 Table 1) and references therein, the MAC/AMS experiments were designed to be as  
751 comparable as possible, i.e. experimental conditions were kept similar between  
752 individual experiments, with light conditions, relative humidity and initial VOC/NO<sub>x</sub>  
753 ratios kept roughly the same from experiment-to-experiment. The reader is guided  
754 to a full discussion of these experiments throughout Sections 2 and 4 (e.g. Alfarra et  
755 al., 2012 and 2013) to ensure that they are fully apprised of the experiments  
756 employed.

757 The main difference between the oxidation systems investigated was in terms of  
758 reactivity of the various precursor compounds with respect to O<sub>3</sub> and OH. Indeed, we  
759 purposefully chose a sufficient span of precursors to cover a range of reactivities and  
760 as such, each system exhibited a different overall reactivity, but on the time scales of  
761 our experiments, the systems were kept as comparable as possible. The references to  
762 our previous work have been made in part to account for/explain this.

763 There was indeed sufficient oxidant throughout all experiments. As described in  
764 Alfarra et al., throughout each of the experiments discussed, O<sub>3</sub> was produced  
765 through system photochemistry. Once formed O<sub>3</sub> was able to react appreciably with  
766 each of the target BVOC precursors. During transit through each respective O<sub>3</sub>-  
767 alkene reaction mechanism and in the presence of water vapor, OH was also  
768 produced. As can be seen in Alfarra et al. (2013), with the exception of β-  
769 caryophyllene, each experiment was characterized by constant net O<sub>3</sub> production  
770 (and hence concomitant OH production), meaning there was always sufficient  
771 oxidant to allow the formation and reaction of second generation products; this can  
772 also be seen in the VOC temporal profiles presented in Alfarra et al., 2013. β-  
773 caryophyllene was observed to consume O<sub>3</sub> as it was produced and net O<sub>3</sub> production  
774 was observed only during the latter stages of the experiments; however during the  
775 net O<sub>3</sub> production stage, there was again sufficient oxidant to allow the formation  
776 and reaction of secondary products, this can be clearly seen in our companion work

777 on  $\beta$ -caryophyllene (Jenkin et al., 2012).  
778 Furthermore, from measurement of the  $m/z$  44 and 43 ions as a function of time in  
779 the AMS data, SOA chemical transformation was observed and fully reported, again  
780 in Alfarra et al., 2013.

781 However for further clarity, the reader has been directed to the work by Hamilton et  
782 al. Jenkin et al. and Alfarra et al. again, following the text describing particle phase  
783 results, in section 4.5, page 1672:

784 “For full details regarding the chemical evolution of the particle phase and the  
785 particle-phase specific experiments in general, see Hamilton et al. (2011), Jenkin et  
786 al., (2012) and Alfarra et al. (2012 and 2013)”.

---

787

788 11) P. 1677, lines 25-27: Can the authors give a statistic that relates the similarity of  
789 the “model mass spectra” to the experimental data set from the chambers? If this  
790 model mass spectra generated from the MCM is similar enough to the experimental  
791 MS obtained, why not run the PCA on MCM generated model mass spectra for each  
792 system and see if the results are similar to the PCA for the gas and particle-phase  
793 data? Or at least, in Discussion section, when trying to relate ions characteristic of  
794 particle-phase oxidation products from grouped precursors, can a metric be used for  
795 testing similarity of the MCM model mass spectra expected and the weight of ions  
796 affecting the various  $\beta$ -caryophyllene and myrcene mappings for example?

797 **Author Response:** A comparison between the model and measured mass spectra is  
798 not straightforward and would first require knowledge of all of the species present in  
799 the gas phase and the exact fragmentation patterns and ionisation routes for each of  
800 these species. Furthermore, one would need to understand all of the isobaric  
801 instances in the measured mass spectra; a difficult task when poorly characterised  
802 ions fragment and overlap. One would then have to (re-) construct mass spectra  
803 representing the gas phase matrices for each of the experiments employed, with no  
804 detection method specific alteration to the mass of each parent compound (e.g.  
805 protonation, adduct formation, fragmentation etc) or instrument biases (a simplified,

806 first approximation attempt at this was made for  $\alpha$ -pinene in Camredon et al., 2010).  
807 As we did with Wyche et al., 2009 and Rickard et al., 2010, we plan to work on a  
808 follow up/companion modelling paper to the current work, in which we  
809 perform/present more complex and in-depth work than is possible to include here,  
810 including the statistical analyses of modelled mass spectra as suggested by the  
811 referee. Unfortunately, to include all of this in one piece of work would make the  
812 product far too large and also divert from the original aims of the paper. The  
813 purpose of the modelling exercise performed here is simply to help us take steps  
814 towards understanding how useful these techniques are for application in a real  
815 world environment, and the discussion presented does just that. What the referee  
816 proposes here is an excellent idea, but a substantial and separate piece of work  
817 entirely. (Note, that the MCM scheme currently does not yet include mechanisms for  
818 myrcene or linalool, and is not directly coupled to partitioning code).

---

819

820 12) P. 1678, lines 26-29: This argument would be strengthened if the authors did  
821 similar MCM analyses as in Fig. 10 for the other precursor types represented in Fig. 9  
822 and provided metrics for the similarity of the model mass spectra with that observed  
823 in the chamber experiments.

824 **Author Response:** As already stated, the purpose of performing the MCM analysis  
825 was to help us move the technique towards a “real-world” environment and to give  
826 some indication as to whether the techniques successfully employed in chamber and  
827 mesocom environments could be useful in the chemically more complex real world  
828 atmosphere. As pointed out by the referees, we are fully open regarding potential  
829 limitations in moving towards the real atmosphere, yet we believe at this stage of our  
830 work, the non-idealized simulations reported, provide a sufficiently chemically  
831 complex model with which to support the arguments made.

832 We feel that performing the simulations using the  $\alpha$ -pinene mechanism was the most  
833 appropriate of the four options available, i.e.  $\alpha$ -pinene, limonene, isoprene and  $\beta$ -  
834 caryophyllene (note, as yet no mechanisms are available in the MCM for myrcene and  
835 linalool oxidation).  $\alpha$ -pinene was chosen primarily because it is the most well

characterised of the available options and moreover both limonene and  $\beta$ -caryophyllene are recent additions to the framework and therefore less well understood. Moreover, to perform the analyses proposed by the referee would be a substantial undertaking and extend the current work too greatly. It is the intention of the authors to work on a follow up/companion modelling paper (as we did with Wyche et al., 2009 and Rickard et al., 2010), where we perform/present more complex and in-depth work than is possible to include here, including statistical analyses on the modelled mass spectra.

As described in point 8), a comparison between the model and measured mass spectra would first require knowledge of all of the species present in the gas-phase and the exact fragmentation patterns and ionisation routes for each of these species. Furthermore, one would need to understand all of the isobaric instances in the measured mass spectra; a difficult task when poorly characterised ions fragment and overlap. However, this is not the current focus of this work, indeed one of the major points of the presented work is to employ the “mass spectral fingerprints” rather than speciated detail.

---

13) P. 1679, line 12: Authors claim that the gas-phase oxidation products of each structural type can be grouped “according to the controlling chemistry and the products formed.” While it is evident that the products formed (as interpreted by the accompanying MS) informs this grouping, what chemical controls are derived from the PCA analysis? I think of e.g. varying NO<sub>x</sub> and O<sub>3</sub> levels as different chemical regimes/controls on the chemistry, but this point is not emphasized in the PCA mapping of the various experiments grouped by precursor. Where is this shown/taken from the figures? Or, please clarify what is meant here.

**Author Response:** The referee is indeed correct, under the current analysis we have not demonstrated this and the text has been amended accordingly. To this end we have included a small separate analysis of the toluene data, which was conducted under a reasonable range of VOC/NO<sub>x</sub> conditions (low, medium, high NO<sub>x</sub>). This new information has been included in a new section, Section 5.3 along with an additional

866 *Figure (10), shown here. Both the additional text and the new Figure 10 were*  
867 *provided following the second comment made by referee 1 (refer to Response to*  
868 *Referee 1).*

---

869

870 14) P. 1679, lines 13-15: In line with comment above, this sentence is rather vague.  
871 No description thus far has really emphasized “decoding of mechanisms” by starting  
872 with an ensemble of summed mass spectra. I agree you can get to the precursor  
873 compounds based on tracer ions indicative of chemistry from a particular precursor,  
874 but what in the PCA analysis maps to particular mechanistic pathways? Do the  
875 authors envision additional mappings in the PCA plots within the precursor groups to  
876 show oxidation from one oxidant versus another, for example, which would be  
877 evident in unique tracer compounds from associated pathways?

878 *Author Response: The referee is correct that this requires some clarification and we*  
879 *believe that the additional section regarding toluene/NO “mapping”, as discussed*  
880 *above under point 14), goes some way towards achieving this.*

881 *In a sense this sentence was pointing to the questions of whether, if you have similar*  
882 *product ions, would the PCA more closely cluster the different mechanistic*  
883 *pathways? The referee is correct in the summation that we envisage additional*  
884 *mappings. The text has been clarified as follows to point to this:*

885 “Indeed, a major potential strength of the data analysis methodology described  
886 here, could lie in the decoding of mechanisms into pathways (i.e. separation within a  
887 group on account of different underlying chemistry) and consequently linking  
888 chemical pathways to precursor compounds.”

---

889

890 *Furthermore, following similar comments made by referee 1, a small additional sub-*  
891 *section (5.2) was added on page 1676, after line 2, (see Reply to Referee 1, first*  
892 *comment):*

893 “5.2 Mapping within a class...”

---

894

895 Technical Comments:

1) P. 1659, Experiment design: It is not clear in Table 1 which experiments were conducted in which chamber.

*Author Response: Table 1 and the accompanying caption have been altered to indicate which experiments were conducted in which chamber.*

---

2) P. 1661, lines 14-16: Since it is not clear which experiments were conducted in which chamber, as in the proceeding comment, please also make clear here which experiments actually had particle-phase data. Later inspection of Figures 7-8 warrants the reader questioning of why the mesocosm experiments and isoprene experiments are not included in the analysis of the particle-phase data.

*Author Response: As Table 1 and the accompanying caption have now been altered to indicate which experiments were conducted in which chamber (as per Referee 2, Technical Comment 1), the reader can now easily see which experiments were conducted in the MAC and therefore which had accompanying particle-phase data. No data was available for certain mesocosm experiments, owing to the low detection limits imposed by the experimental set up and for those experiments with particle-phase data, insufficient repeat experiments were available to use in the statistical analysis presented here.*

---

3) Figure 2 would benefit from caption description of the mass spectral ions populating the figure.

*Author Response: Figure 2 caption has been altered accordingly to include a description of the major mass spectral ions populating the figure.*

---

4) P. 1669, line 13: Change “been” to “be”

*Author Response: The text has been altered accordingly.*

---

5) For aid in interpretation of the ions mapped onto figures 2 and 7a, it may be

924 helpful to include a table that lists all ions (LC-MS/MS, CIR-Tof-MS, and AMS), their  
925 chemical assignment, and precursor type, so the information is more readily  
926 available than filtering through the text for these ion assignments.

927 *Author Response: The requested table has been produced (now Table 3) and inserted*  
928 *and the relevant main text and captions have been altered accordingly.*

---

929

930 6) Figures 7b and 8b: It would be more helpful if the color scheme used for the cyclic  
931 monoterpenes was consistent with that used in Figure 5.

932 *Author Response: We thank the referee for pointing this out, but we would prefer to*  
933 *leave the dendrogram colors as there are, as they represent the best combination for*  
934 *the clearest description of the data, i.e. the dendrogram colors were chosen*  
935 *specifically to optimize clarity of view within the figure.*

---

936



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

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,  $\alpha$ -pinene, limonene, myrcene, linalool and  $\beta$ -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 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

999 | sources and controlling oxidative chemistry. The potential to extend the  
1000 methodology to the analysis of ambient air is discussed using results obtained from a  
1001 zero-dimensional box model incorporating mechanistic data obtained from the  
1002 Master Chemical Mechanism (MCMv3.2). Such an extension to analysing ambient  
1003 air would prove a powerful asset in assisting with the identification of SOA sources  
1004 and the elucidation of the underlying chemical mechanisms involved.

1005  
1006  
1007 *Keywords: Volatile organic compounds, secondary organic aerosol, environmental*  
1008 *simulation chamber, photooxidation, principal component analysis, cluster analysis,*  
1009 *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<sup>-1</sup> (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<sub>5</sub>), monoterpenes (C<sub>10</sub>) and sesquiterpenes (C<sub>15</sub>) (e.g. (Atkinson and Arey, 2003a;Kanakidou et al., 2005)).

Within the troposphere terpenes are able to react with OH, O<sub>3</sub> and NO<sub>3</sub> 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

1043 human health, particularly in vulnerable members of the population (Harrison et al.,  
1044 2010;Heal et al., 2012).

1045

1046 Biogenic SOA (BSOA) has been estimated to account for a significant fraction of total  
1047 global SOA. Modelling studies suggest the annual global production rate of BSOA is  
1048 of the order 16.4 Tg Yr<sup>-1</sup> (Henze and Seinfeld, 2006). However, despite its importance  
1049 and the significant amount of investigation conducted upon it, the formation  
1050 mechanisms and chemical composition of BSOA are still not well characterised (e.g.  
1051 (Librando and Tringali, 2005;Wang et al., 2013). Indeed under certain conditions as  
1052 much as 80 – 90 % of analysed SOA mass is unknown (Limbeck et al., 2003;Kalberer  
1053 et al., 2006). In particular, there remains a significant lack of information regarding  
1054 the composition and evolution of the complex organic gas-phase matrix during  
1055 aerosol formation, and its linkage to SOA (Kroll et al., 2005;Librando and Tringali,  
1056 2005). Indeed, in the many studies conducted on BSOA, very few oxidation products  
1057 of the precursor are routinely identified and reported.

1058

1059 The chemistry of the atmospheric system is highly non-linear and can be studied by  
1060 experiments ranging from highly controlled laboratory studies of a single process, to  
1061 field studies of the whole complex system. A significant proportion of the findings  
1062 gained regarding SOA over the last decade and more have come from atmospheric  
1063 simulation chamber experiments, intermediate in complexity between classical  
1064 single-process experiments and the fully open system (for various different chamber  
1065 systems and VOC precursors, see for example, Pandis et al., 1991; Odum et al., 1996;  
1066 Hoffmann et al., 1997; Griffin et al., 1999; Glasius et al., 2000; Cocker et al., 2001;  
1067 Jaoui and Kamens, 2003; Kleindienst et al., 2004; Presto et al., 2005; Bloss et al.,  
1068 2005; Rohrer et al., 2005; Ng et al., 2006, 2007; Dommen et al., 2006; Surrat et al.,  
1069 2006; Grieshop et al., 2007; Chan et al., 2007; Wyche et al., 2009; Hildebrandt et al.,  
1070 2009; Rickard et al., 2010; Camredon et al., 2010; Chhabra et al., 2011; Hennigan et  
1071 al., 2011; Jenkin et al., 2012). Chamber experiments produce a large amount of data,  
1072 the interpretation of which can often be highly complex and time consuming even  
1073 though the set-up of the chamber constrains the complexity to a large degree.

1074

Kevin Wyche 15/4/2015 15:57

**Deleted:** Atmospheric

Kevin Wyche 15/4/2015 15:57

**Deleted:** is a

Kevin Wyche 15/4/2015 15:58

**Deleted:** system which

Kevin Wyche 9/5/2015 13:30

**Deleted:** A significant proportion of the findings gained regarding SOA over the last decade and more have come from atmospheric simulation chamber experiments (e.g. (Jenkin et al., 2012;Wyche et al., 2009;Rickard et al., 2010;Camredon et al., 2010)), intermediate in complexity between classical single-process experiments and the fully open system

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  $1.1 \times 10^7$  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.,

Kevin Wyche 15/4/2015 15:59

Deleted: ,

Kevin Wyche 15/4/2015 15:59

Deleted: chemical ionisation reaction time-of-flight mass spectrometry (

Kevin Wyche 15/4/2015 15:59

Deleted: )

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  $\beta$ -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.

Kevin Wyche 6/5/2015 14:16

**Deleted:** However our approach investigates both the gas- and particle-phases and also provides insight into the fundamental chemical reaction pathways. -



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<sub>3</sub> 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<sub>5</sub> diene that accounts for around 62 % (~ 594 Tg yr<sup>-1</sup>) of total annual non-methane BVOC emissions (Sindelarova et al., 2014). After isoprene, monoterpenes (C<sub>10</sub>H<sub>16</sub>) have the next largest annual emission rate, they account for around 11 % (~ 95 Tg yr<sup>-1</sup>) 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

Kevin Wyche 9/5/2015 13:33

**Deleted:** (e.g. (Lee et al., 2006;Alfarra et al., 2013) and references therein)

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,  $\beta$ -caryophyllene was included to represent sesquiterpenes, which have annual emissions of the order 20 Tg yr<sup>-1</sup> (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<sub>3</sub>, e.g. limonene, myrcene, linalool and  $\beta$ -caryophyllene all have atmospheric lifetimes with respect to OH of the order 40 – 50 minutes (Alfarra et al., 2013; Atkinson and Arey, 2003b).  $\beta$ -caryophyllene has the shortest lifetime with respect to O<sub>3</sub> (ca. 2 minutes) and isoprene and  $\alpha$ -pinene have the longest lifetimes with respect to both OH and O<sub>3</sub>, e.g. isoprene and  $\alpha$ -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<sub>2</sub> (VOC/NO<sub>x</sub> 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 ( $\Sigma$ VOC, *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 ( $T = 40$  °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 ( $H_3O^+ \cdot (H_2O)_n$ ) 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 °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<sub>2</sub>, NO<sub>x</sub> analysers, UV photometric O<sub>3</sub> 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<sub>x</sub> ratio (Alfarra et al., 2013)).

## 2.5 Model construction

Kevin Wyche 15/4/2015 16:11

Deleted: (/

Kevin Wyche 15/4/2015 16:11

Deleted: )

In order to aid analysis, the composition and evolution of the gas-phase components of the  $\alpha$ -pinene chamber system were simulated using a chamber optimised photochemical box model incorporating the comprehensive  $\alpha$ -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  $\alpha$ -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  $j(\text{NO}_2)$ . All simulations were run at 295 K and 50 % relative humidity. NO, NO<sub>2</sub>, HONO and  $\alpha$ -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.*  $\Sigma(\text{H}_3\text{O}^+ + \text{H}_3\text{O}^+(\text{H}_2\text{O})_n)$ ). 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 =  $\alpha$ -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 =  $\beta$ -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.  $\beta$ -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  $\alpha$ -pinene (grouped into one class), and the straight chain monoterpene experiments of

Kevin Wyche 9/5/2015 13:36

**Deleted:** ((Bloss et al., 2005;Wyche et al., 2009;Camredon et al., 2010;Rickard et al., 2010;Hamilton et al., 2011;Jenkin et al., 2012;Alfarra et al., 2012;Alfarra et al., 2013;Wyche et al., 2014) and references therein)



1454 myrcene and linalool (grouped into a second class), albeit with the latter having a  
1455 greater spread in confidence.

1456

1457 The  $m/z$  loadings of the PCA allow us to understand how the spectral fingerprints of  
1458 the different terpene oxidation systems are grouped/separated by the PCA model.

1459 The first set of ions that contribute to separation of the different terpene systems  
1460 comprises the protonated parent ions ( $MH^+$ ) of the precursors themselves (and  
1461 major fragments thereof), *i.e.*  $m/z$  69 for isoprene, 137 (and fragment 81) for all  
1462 monoterpenes (regardless of structure) and 205 for  $\beta$ -caryophyllene. Important  
1463 contributions are to be expected from the respective parent-ions (being the basis for  
1464 the use of chemical-ionisation mass spectrometry as an analyser of gas mixtures  
1465 (Blake et al., 2009)). Our purpose here goes beyond identification of precursor and

1466 intermediate VOCs to an interpretation of reaction pathways in complex mixtures  
1467 and potential linkages to SOA. In doing this, a certain amount of disambiguation of  
1468 isobaric compounds becomes possible; indeed, as discussed in more detail below,  
1469 Figure 2 clearly shows separation between cyclic and non-cyclic monoterpene  
1470 oxidation groups, both of which have precursors of molecular weight (MW) 136 g  
1471  $\text{mol}^{-1}$ . Note, for clarity within Figure 2, the scale has been set to show the bulk of the  
1472 data, hence precursor parent ions and  $m/z$  71 are not shown.

1473

1474 Moving past the precursors into the detailed chemical information provided by the  
1475 oxidation products formed within the chamber, we can see from Figure 2 that  
1476 amongst others,  $m/z$  71 (methyl vinyl ketone and methacrolein), 75 (hydroxy  
1477 acetone), 83 (methyl furan) and 87 ( $C_4$ -hydroxycarbonyls/methacrylic acid) all  
1478 contribute to separation of the isoprene group, and  $m/z$  237 ( $\beta$ -caryophyllene  
1479 aldehyde) and 235 and 253 ( $\beta$ -caryophyllene secondary ozonide and isomers  
1480 thereof) to that of the  $\beta$ -caryophyllene group. The monoterpene groupings are  
1481 influenced by the presence of  $m/z$  107, 151 and 169 (primary aldehydes-  
1482 pionaldehyde and limonondehyde) and 139 (primary ketone- limonaketone)  
1483 ions in their mass spectra. Helping to separate the straight chain from cyclic  
1484 monoterpenes are  $m/z$  95 and 93, relatively dominant features in both the myrcene  
1485 and linalool spectra (relative abundance 10 – 24 % for  $m/z$  93).  $m/z$  93 has

Kevin Wyche 15/4/2015 16:13

Deleted: caryophyllon

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

Kevin Wyche 8/5/2015 14:02

**Deleted:** 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.

(*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

Kevin Wyche 15/4/2015 16:18

**Deleted:** As a further test of the technique to distinguish between and to classify VOCs and their oxidized atmospheres, test data from an anthropogenic system was introduced into the model.

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 =  $\alpha$ -pinene and limonene; sesquiterpene =  $\beta$ -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

Kevin Wyche 15/4/2015 16:28

**Deleted:** The greater spread in confidence of the noncyclic monoterpene group is once again most likely owing to the low number of repeats employed for only two types of precursor.

Kevin Wyche 16/4/2015 13:01

**Deleted:** en

Kevin Wyche 16/4/2015 15:39

**Deleted:** 3

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 ( $\beta$ -caryophyllonic acid), 255 (4-(2-(2-carboxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid), 267 ( $\beta$ -14-hydroxycaryophyllonic acid and  $\beta$ -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  $\beta$ -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,  $\beta$ -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 limonic acid), 183 (pinonic acid, limonic 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  $\alpha$ -pinene experiments). For further details regarding  $\alpha$ -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. \(2006\),](#)

Kevin Wyche 9/5/2015 13:37

**Deleted:** see Hamilton et al., (Hamilton et al., 2011) and Jenkin et al. (Jenkin et al., 2012) and the Section 5

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  $[M-H_2+FA+Na]^+$   $M = 254$  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.*  $\beta$ -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,  $\alpha$ -pinene, limonene and linalool tend in general to cluster towards the upper and right regions of the PCA space, primarily

Kevin Wyche 9/5/2015 13:38

**Deleted:** Jenkin et al. (Jenkin, 2004), Lee et al. (Lee et al., 2006), Camredon et al., (Camredon et al., 2010) and Hamilton et al. (Hamilton et al., 2011)

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  $\text{CH}_3\text{CO}^+$  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 ( $\beta$ -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  $\text{C}_3\text{H}_5^+$  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  $\alpha$ -pinene and  $\beta$ -caryophyllene clusters in the PCA space) is also influenced by  $m/z$  44, *i.e.* most likely the  $\text{CO}_2^+$  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 dicarboxylic 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  $\text{NO}_x$ , *e.g.*  $\text{C}_4$ -hydroxy nitrate peroxy acetyl

nitrate (C4-HN-PAN in Figure 9) (Surratt et al., 2010), as well as condensable OVOCs, *e.g.* hydroxymethyl-methyl- $\alpha$ -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<sub>x</sub>” 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<sub>4</sub> and C<sub>5</sub> saturated, low volatility species constitute the monomer building blocks that proceed to form relatively high O:C ratio (nitrated in the presence of NO<sub>x</sub> and sulphated in the presence of H<sub>2</sub>SO<sub>4</sub>) 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<sup>+</sup> = *m/z* 71), isoprene nitrates (ISOPN in Figure 9; MH<sup>+</sup> - HNO<sub>3</sub> = *m/z* 85) and MPAN (MH<sup>+</sup>.H<sub>2</sub>O - HNO<sub>3</sub> = *m/z* 103) under “high NO<sub>x</sub>” 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<sup>+</sup> - H<sub>2</sub>O = *m/z* 101) under “low NO<sub>x</sub>” conditions. For the “high NO<sub>x</sub>” 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<sub>4</sub>-hydroxycarbonyls/methacrylic acid, ISOPN, C<sub>5</sub>-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<sup>+</sup> = *m/z* 103) may produce fragment ions of *m/z* 85 (*i.e.* MH<sup>+</sup>-H<sub>2</sub>O) 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<sub>3</sub>), producing gas-phase spectra with mid MW C<sub>9</sub> and C<sub>10</sub> oxygenated (and nitrated in the presence of NO<sub>x</sub>) 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 dimer, 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<sub>7</sub> OVOCs (and/or NVOCs, depending on NO<sub>x</sub> levels). Reaction at the conjugated double bond system in myrcene would be expected to form formaldehyde in conjunction with either a C<sub>9</sub> aldehyde or C<sub>9</sub> 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  $1 \times 10^{-10}$  (*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<sub>9</sub> and C<sub>7</sub> 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<sup>+</sup> and MH<sup>+</sup>-H<sub>2</sub>O, respectively) and 113 and 95 (*e.g.* 2-methylenepentanedial MH<sup>+</sup> and MH<sup>+</sup>-H<sub>2</sub>O, 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<sub>9</sub> aldehyde and/or C<sub>9</sub> ketone product) also assists in separating the myrcene spectra from those of α-pinene.

1766

1767 By comparing both the gas- and particle-phase cyclic monoterpenes in Figures 2 and  
1768 7a, it is evident that the dominant loadings represent compounds of similar MW, *i.e.*  
1769 169, 151 and 107 (primary aldehyde product, *e.g.* pinonaldehyde- PINAL in Figure 9,  
1770 parent ion and fragments thereof) and 139 (primary ketone product parent ion) for  
1771 the gas-phase and 187, 185, 183 and 169 for the particle-phase. Conversely, for the  
1772 straight chain monoterpene experiments the major gas-phase loadings represent  
1773 compounds of significantly smaller MW than their particle-phase counterparts, *i.e.*  
1774 113 and 95 and 111 and 93, compared to 325, 322, 321, 227 and 215. Indeed, the  
1775 straight chain monoterpene LC-MS/MS spectra contained on average ~ 10 % more  
1776 signal > 250 Da than the cyclic monoterpene spectra. Also, the composition of the  
1777 ions observed in the straight chain monoterpene LC-MS/MS spectra suggests that  
1778 the SOA particles contained both oligomers and highly oxidized species, with the C<sub>10</sub>  
1779 backbone intact (*i.e.* O:C = 0.6), similar in structure to (but a little less oxidised than)  
1780 extremely low volatility organic vapours (ELV-VOC), which have been observed  
1781 previously in significant yield from  $\alpha$ -pinene and limonene (as well as 6-nonenal)  
1782 ozonolysis chamber experiments in the absence of an OH scavenger, as well as  
1783 boreal forests in Finland (Ehn et al., 2014). Further evidence to elucidate the type  
1784 of SOA formed from the oxidation of straight chain monoterpenes can be obtained  
1785 from investigation of the grouping of myrcene spectra in the cTOF-AMS PCA (Figure  
1786 8a). In the hour-4 cTOF-AMS PCA loadings bi-plot, we see that the grouping of the  
1787 myrcene spectra is influenced somewhat by both *m/z* 41 and 44, indicating the  
1788 presence of LV-OOA in the SOA, potentially a result of oligomerisation or further  
1789 oxidative heterogeneous chemistry involving reaction at remaining C=C double bond  
1790 sites.

1791

1792  $\beta$ -caryophyllene readily forms particulate matter on oxidation ([e.g. Jaoui et al., 2003;](#)  
1793 [Lee et al., 2006; Winterhalter et al., 2009; Alfarra et al., 2012; Chen et al., 2012](#)), with  
1794 reaction predominantly at one of the two C=C sites (*e.g.* with OH or O<sub>3</sub>, although O<sub>3</sub>  
1795 attack occurs almost exclusively at the endocyclic double bond (Jenkin et al., 2012)),  
1796 yielding relatively low vapour pressure, unsaturated and oxygenated primary  
1797 products (Figure 9), which have significant affinity for the particle-phase (Jenkin et

Kevin Wyche 9/5/2015 13:40

Deleted: (Alfarra et al., 2012)

al., 2012). A further oxidation step involving the second C=C site can result in increased oxygen (and/or nitrogen, depending on NO<sub>x</sub> 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  $\beta$ -caryophyllene oxidation and SOA formation with comparable MW species, *e.g.* primary products  $\beta$ -caryophyllon aldehyde (MW 236, BCAL in Figure 9) and  $\beta$ -caryophyllene secondary ozonide in the gas-phase (MW 252, BCSOZ in Figure 9),  $\beta$ -caryophyllonic acid (MW 252, C141CO<sub>2</sub>H in Figure 9) in both phases and secondary product  $\beta$ -nocaryophyllinic acid (MW 254, C131CO<sub>2</sub>H in Figure 9) in the particle-phase. In the hour-4 cTOF-AMS PCA scores plot, the myrcene and  $\beta$ -caryophyllene clusters are located adjacent to one another, with  $\beta$ -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  $\alpha$ -pinene experiments, with three out of the four  $\alpha$ -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,  $\alpha$ -pinene and limonene react at somewhat different rates with respect to both OH and O<sub>3</sub> (Atkinson and Arey, 2003). 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  $\alpha$ -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<sub>x</sub> ratios. The VOC/NO<sub>x</sub> ratios employed nominally represent “low”, “medium” and “high” NO<sub>x</sub> conditions, with values of roughly 11 (*i.e.* low NO<sub>x</sub>, “NO<sub>x</sub>-limited” ozone formation conditions – as determined from simulation chamber ozone isopleth plots, see Wagner et al., 2003; two experiments), 4 (*i.e.* moderate NO<sub>x</sub>, two experiments) and 1 (*i.e.* high NO<sub>x</sub>, “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<sub>x</sub> ratios. Figure 10 shows the low NO<sub>x</sub>, high VOC/NO<sub>x</sub> 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<sub>x</sub> levels in the NO<sub>x</sub> limited regime, result in low [OH] (reduced radical cycling) and low [O<sub>3</sub>] (less NO to NO<sub>2</sub> conversions) (see also Bloss et al., 2005). Similarly, the relatively large contribution from cresol to the low NO<sub>x</sub> 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<sub>x</sub>, medium VOC/NO<sub>x</sub> 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  $\text{NO}_x$  conditions. Larger net benzaldehyde concentrations originate from greater system reactivity and greater abundance of NO to fuel the  $\text{RO}_2 + \text{NO}$  reaction.

The high  $\text{NO}_x$ , low VOC/ $\text{NO}_x$  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  $\text{NO}_x$  experiments) and the resultant greater proportion of ring-opening product ions (as the case for the moderate  $\text{NO}_x$  experiments). However, the low VOC/ $\text{NO}_x$  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  $\text{NO}_x$  conditions, owing to reaction through the  $\text{RO}_2 + \text{NO}$  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  $\alpha$ -pinene system and operated under three different scenarios:

- (1) *Basic chamber simulation*:  $\alpha$ -pinene concentration constrained to measurements (initial concentration 124 ppbV); NO and NO<sub>2</sub> initialised according to measurements (31 and 41 ppbV, respectively).
- (2) *Spiked chamber simulation*:  $\alpha$ -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*:  $\alpha$ -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 \( \$\alpha\$ -pinene\) and Jenkin et al., 2012 \( \$\beta\$ -caryophyllene\).](#)

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  $\alpha$ -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 ( $\Delta$ MC) is generally  $\leq 2\%$ , with the exceptions of MWs 168 and 186. MW 168 primarily comprises pinonaldehyde, with a  $\Delta$ MC 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  $\Delta$ MC of roughly  $+3\%$ .

Kevin Wyche 7/5/2015 14:57  
Deleted: 10

Kevin Wyche 7/5/2015 14:58  
Deleted: 10a

Kevin Wyche 7/5/2015 14:58  
Deleted: 10a

Kevin Wyche 7/5/2015 14:58  
Deleted: 10b

Kevin Wyche 7/5/2015 14:58  
Deleted: 10a

Kevin Wyche 7/5/2015 14:58  
Deleted: 10c

Kevin Wyche 7/5/2015 14:58  
Deleted: 10c

Kevin Wyche 7/5/2015 14:58  
Deleted: 10d

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  $\Delta MC$  is generally  $\leq \pm 5\%$ , 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 major potential 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

Kevin Wyche 8/5/2015 12:17

**Deleted:** Indeed, a major strength of the data analysis methodology described here, lies in the decoding of mechanisms into pathways and consequently linking the pathways to precursor compounds

2032 were supported by the NERC National Centre for Atmospheric Sciences (NCAS). The  
2033 authors would like to thank the University of Leicester Atmospheric Chemistry group  
2034 for assistance throughout all experiments, including Drs Alex Parker, Chris Whyte,  
2035 Iain White and Timo Carr; co-workers at the University of Manchester for assistance  
2036 with MAC experiments; co-workers from Fundacion CEAM, Drs Marie Camredon and  
2037 Salim Alam for assistance with EUPHORE experiments and co-workers from the  
2038 Laboratory of Atmospheric Chemistry smog chamber facility at the Paul Scherrer  
2039 Institute (PSI) for assistance with PSISC experiments.

2040

## 2041 9. References

2042 Alfarra, M. R., Coe, H., Allan, J. D., Bower, K. N., Boudries, H., Canagaratna, M. R.,  
2043 Jimenez, J. L., Jayne, J. T., Garforth, A. A., Li, S. M., and Worsnop, D. R.:  
2044 Characterization of urban and rural organic particulate in the lower Fraser valley  
2045 using two aerodyne aerosol mass spectrometers, *Atmospheric Environment*, 38,  
2046 5745-5758, 10.1016/j.atmosenv.2004.01.054, 2004.

2047 Alfarra, M. R., Paulsen, D., Gysel, M., Garforth, A. A., Dommen, J., Prevot, A. S. H.,  
2048 Worsnop, D. R., Baltensperger, U., and Coe, H.: A mass spectrometric study of  
2049 secondary organic aerosols formed from the photooxidation of anthropogenic and  
2050 biogenic precursors in a reaction chamber, *Atmospheric Chemistry and Physics*, 6,  
2051 5279-5293, 2006.

2052 Alfarra, M. R., Hamilton, J. F., Wyche, K. P., Good, N., Ward, M. W., Carr, T., Barley,  
2053 M. H., Monks, P. S., Jenkin, M. E., Lewis, A. C., and McFiggans, G. B.: The effect of  
2054 photochemical ageing and initial precursor concentration on the composition and  
2055 hygroscopic properties of beta-caryophyllene secondary organic aerosol,  
2056 *Atmospheric Chemistry and Physics*, 12, 6417-6436, 10.5194/acp-12-6417-2012,  
2057 2012.

2058 Alfarra, M. R., Good, N., Wyche, K. P., Hamilton, J. E., Monks, P. S., Lewis, A. C., and  
2059 McFiggans, G.: Water uptake is independent of the inferred composition of  
2060 secondary aerosols derived from multiple biogenic VOCs, *Atmospheric Chemistry  
2061 and Physics*, 13, 11769-11789, 10.5194/acp-13-11769-2013, 2013.

2062 Alier, M., van Drooge, B. L., Dall'Osto, M., Querol, X., Grimalt, J. O., and Tauler, R.:  
2063 Source apportionment of submicron organic aerosol at an urban background and a  
2064 road site in Barcelona (Spain) during SAPUSS, *Atmospheric Chemistry and Physics*,  
2065 13, 10353-10371, doi:10.5194/acp-13-10353-2013, 2013.

2066 Atkinson, R., and Arey: Gas phase tropospheric chemistry of biogenic volatile organic  
2067 compounds- a review, *Atmospheric Environment*, 37, S197 - S219, 2003a.

- 2068 Atkinson, R., and Arey, J.: Atmospheric Degradation of Volatile Organic Compounds,  
2069 Chem. Rev., 103, 4605-4638, 2003b.
- 2070 Baltensperger, U., Dommen, J., Alfarra, R., Duplissy, J., Gaeggeler, K., Metzger, A.,  
2071 Facchini, M. C., Decesari, S., Finessi, E., Reinnig, C., Schott, M., Warnke, J., Hoffmann,  
2072 T., Klatzer, B., Puxbaum, H., Geiser, M., Savi, M., Lang, D., Kalberer, M., and Geiser,  
2073 T.: Combined determination of the chemical composition and of health effects of  
2074 secondary organic aerosols: The POLYSOA project, Journal of Aerosol Medicine and  
2075 Pulmonary Drug Delivery, 21, 145-154, 10.1089/jamp.2007.0655, 2008.
- 2076 Becker, K. H.: The European Photoreactor EUPHORE, Final Re-port to the European  
2077 Commission, Bergische Universitat Wuppertal, Germany, Wuppertal, 1996.
- 2078 Blake, R. S., Whyte, C., Hughes, C. O., Ellis, A. M., and Monks, P. S.: Demonstration of  
2079 proton-transfer reaction time-of-flight mass spectrometry for real-time analysis of  
2080 trace volatile organic compounds, Analytical Chemistry, 76, 3841-3845,  
2081 10.1021/ac0498260, 2004.
- 2082 Blake, R. S., Wyche, K. P., Ellis, A. M., and Monks, P. S.: Chemical ionization reaction  
2083 time-of-flight mass spectrometry: Multi-reagent analysis for determination of trace  
2084 gas composition, International Journal Of Mass Spectrometry, 254, 85-93, 2006.
- 2085 Blake, R. S., Monks, P. S., and Ellis, A. M.: Proton Transfer Reaction Mass  
2086 Spectrometry, Chem. Rev., 109, 861-896, 2009.
- 2087 Bloss, C., Wagner, V., Bonzanini, A., Jenkin, M. E., Wirtz, K., Martin-Reviejo, M., and  
2088 Pilling, M. J.: Evaluation of detailed aromatic mechanisms (MCMv3 and MCMv3.1)  
2089 against environmental chamber data, Atmospheric Chemistry and Physics 5, 623-  
2090 639, 2005.
- 2091 Bouvier-Brown, N. C., Goldstein, A. H., Gilman, J. B., Kuster, W. C., and de Gouw, J.  
2092 A.: In-situ ambient quantification of monoterpenes, sesquiterpenes, and related  
2093 oxygenated compounds during BEARPEX 2007: implications for gas- and particle-  
2094 phase chemistry, Atmospheric Chemistry and Physics, 9, 5505-5518,  
2095 doi:10.5194/acp-9-5505-2009, 2009.
- 2096 Calogirou, A., Larsen, B. R., and Kotzias, D.: Gas-phase terpene oxidation products: a  
2097 review, Atmospheric Environment, 33, 1423-1439, 10.1016/s1352-2310(98)00277-5,  
2098 1999.
- 2099 Calvert, J. G., Atkinson, R., Kerr, J. A., Madronich, S., Moortgat, G. K., Wallington, T.  
2100 J., and Yarwood, G.: The Mechanisms of Atmospheric Oxidation of the Alkenes,  
2101 Oxford University Press, New York, 2000.
- 2102 Camredon, M., Hamilton, J. F., Alam, M. S., Wyche, K. P., Carr, T., White, I. R., Monks,  
2103 P. S., Rickard, A. R., and Bloss, W. J.: Distribution of gaseous and particulate organic  
2104 composition during dark alpha-pinene ozonolysis, Atmospheric Chemistry and  
2105 Physics, 10, 2893-2917, 2010.

2106 Canagaratna, M. R., Jayne, J. T., Jimenez, J. L., Allan, J. D., Alfarra, M. R., Zhang, Q.,  
 2107 Onasch, T. B., Drewnick, F., Coe, H., Middlebrook, A., Delia, A., Williams, L. R.,  
 2108 Trimborn, A. M., Northway, M. J., DeCarlo, P. F., Kolb, C. E., Davidovits, P., and  
 2109 Worsnop, D. R.: Chemical and microphysical characterization of ambient aerosols  
 2110 with the aerodyne aerosol mass spectrometer, *Mass Spectrometry Reviews*, 26, 185-  
 2111 222, :10.1002/mas.20115, 2007.

2112 Capouet, M., Peeters, J., Noziere, B., and Muller, J. F.: Alpha-pinene oxidation by OH:  
 2113 simulations of laboratory experiments, *Atmospheric Chemistry and Physics*, 4, 2285 -  
 2114 2311, 2004.

2115 [Chan, A. W. H., Kroll, J. H., Ng, N. L., and Seinfeld, J. H.: Kinetic modeling of secondary](#)  
 2116 [organic aerosol formation: effects of particle- and gas-phase reactions of](#)  
 2117 [semivolatile products, \*Atmos. Chem. Phys.\*, 7, 4135–4147, doi:10.5194/acp-7-4135-](#)  
 2118 [2007, 2007.](#)

2119 [Chan, M. N., Surratt, J. D., Chan, A. W. H., Schilling, K., Offenberg, J. H., Lewandowski,](#)  
 2120 [M., Edney, E. O., Kleindienst, T. E., Jaoui, M., Edgerton, E. S., Tanner, R. L., Shaw, S. L.,](#)  
 2121 [Zheng, M., Knipping, E. M., and Seinfeld, J. H.: Influence of aerosol acidity on the](#)  
 2122 [chemical composition of secondary organic aerosol from  \$\beta\$ -caryophyllene. \*Atmos.\*](#)  
 2123 [Chem. Phys., 11, 1735–1751, doi:10.5194/acp-11-1735-2011, 2011.](#)

2124 [Chen, Q., Li, Y. L., McKinney, K. A., Kuwata, M., and Martin, S. T.: Particle mass yield](#)  
 2125 [from  \$\beta\$ -caryophyllene ozonolysis, \*Atmos. Chem. Phys.\*, 12, 3165-3179, doi:](#)  
 2126 [10.5194/acp-12-3165-2012, 2012.](#)

2127 [Chhabra, P. S., Ng, N. L., Canagaratna, M. R., Corrigan, A. L., Russell, L. M., Worsnop,](#)  
 2128 [D. R., Flagan, R. C., and Seinfeld, J. H.: Elemental composition and oxidation of](#)  
 2129 [chamber organic aerosol, \*Atmos. Chem. Phys.\*, 11, 8827–8845, doi:10.5194/acp-11-](#)  
 2130 [8827-2011, 2011.](#)

2131 Claeys, M., Graham, B., Vas, G., Wang, W., Vermeylen, R., Pashynska, V., Cafmeyer,  
 2132 J., Guyon, P., Andreae, M. O., Artaxo, P., and Maenhaut, W.: Formation of secondary  
 2133 organic aerosols through photooxidation of isoprene, *Science*, 303, 1173-1176, 2004.

2134 [Cocker III, D. R., Mader, B. T., Kalberer, M., Flagan, R. C., and Seinfeld, J. H.: The](#)  
 2135 [effect of water on gas-particle partitioning of secondary organic aerosol: II. m-xylene](#)  
 2136 [and 1,3,5-trimethylbenzene photooxidation systems, \*Atmos. Env.\*, 35, 6073-6085,](#)  
 2137 [2001.](#)

2138 Derwent, R. G., Jenkin, M. E., Passant, N. R., and Pilling, M. J.: Photochemical ozone  
 2139 creation potentials (POCPs) for different emission sources of organic compounds  
 2140 under European conditions estimated with Master Chemical Mechanism,  
 2141 *Atmospheric Environment*, 41, 2570-2579, 2007.

2142 [Dommen, J., Metzger, A., Duplissy, J., Kalberer, M., Alfarra, M. R., Gascho, A.,](#)  
 2143 [Weingartner, E., Prevot, A. S. H., Verheggen, B., and Baltensperger, U.: Laboratory](#)  
 2144 [observation of oligomers in the aerosol from isoprene/NO<sub>x</sub> photooxidation,](#)

- 2145 | [Geophys. Res. Lett., 33, L13805, doi:10.1029/2006gl026523, 2006.](#)
- 2146 Drewnick, F., Hings, S. S., DeCarlo, P., Jayne, J. T., Gonin, M., Fuhrer, K., Weimer, S.,  
2147 Jimenez, J. L., Demerjian, K. L., Borrmann, S., and Worsnop, D. R.: A new time-of-  
2148 flight aerosol mass spectrometer (TOF-AMS) - Instrument description and first field  
2149 deployment, *Aerosol Sci. Technol.*, 39, 637-658, 10.1080/02786820500182040,  
2150 2005.
- 2151 Eddingsaas, N. C., Loza, C. L., Yee, L. D., Seinfeld, J. H., and Wennberg, P. O.: alpha-  
2152 pinene photooxidation under controlled chemical conditions - Part 1: Gas-phase  
2153 composition in low- and high-NO<sub>x</sub> environments, *Atmospheric Chemistry and*  
2154 *Physics*, 12, 6489-6504, 10.5194/acp-12-6489-2012, 2012a.
- 2155 Eddingsaas, N. C., Loza, C. L., Yee, L. D., Chan, M., Schilling, K. A., Chhabra, P. S.,  
2156 Seinfeld, J. H., and Wennberg, P. O.: alpha-pinene photooxidation under controlled  
2157 chemical conditions - Part 2: SOA yield and composition in low- and high-NO<sub>x</sub>  
2158 environments, *Atmospheric Chemistry and Physics*, 12, 7413-7427, 10.5194/acp-12-  
2159 7413-2012, 2012b.
- 2160 Ehn, M., Thornton, J. A., Kleist, E., Sipila, M., Junninen, H., Pullinen, I., Springer, M.,  
2161 Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I. H., Rissanen, M.,  
2162 Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurten,  
2163 T., Nielsen, L. B., Jorgensen, S., Kjaergaard, H. G., Canagaratna, M., Dal Maso, M.,  
2164 Berndt, T., Petaja, T., Wahner, A., Kerminen, V. M., Kulmala, M., Worsnop, D. R.,  
2165 Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol,  
2166 *Nature*, 506, 476+, 10.1038/nature13032, 2014.
- 2167 Fantechi, G., Vereecken, L., and Peeters, J.: The OH initiated atmospheric oxidation  
2168 of pinonaldehyde: Detailed theoretical study and mechanism construction, *Physical*  
2169 *Chemistry Chemical Physics*, 4, 5795 - 5805, 2002.
- 2170 [Frey, J. L., Draper, D. C., Barsanti, K. C., Smith, J. N., Ortega, J., P. M., Winkler, Lawler,](#)  
2171 [M. J., Brown, S. S., Edwards, P. M., Cohen, R. C., and Lee, L.: Secondary Organic](#)  
2172 [Aerosol Formation and Organic Nitrate Yield from NO<sub>3</sub> Oxidation of Biogenic](#)  
2173 [Hydrocarbons, Environmental Science & Technology, 48\(20\), 11944-11953,](#)  
2174 [10.1021/es502204x, 2014](#)
- 2175 Gao, S., Ng, N. L., Keywood, M., Varutbangkul, V., Bahreini, R., Nenes, A., He, J., Yoo,  
2176 K. Y., Beauchamp, J. L., Hodyss, R. P., Flagan, R. C., and Seinfeld, J. H.: Particle Phase  
2177 Acidity and Oligomer Formation in Secondary Organic Aerosol, *Environmental*  
2178 *Science & Technology*, 38, 6582-6589, 2004.
- 2179 [Glasius, M., Lahaniati, M., Calogirou, A., Bella, D. D., Jensen, N. R., Hjorth, J., Kotzias,](#)  
2180 [D., and Larsen, B. R.: Carboxylic acids in secondary aerosol from oxidation of cyclic](#)  
2181 [monoterpenes by ozone, Environ. Sci. Technol., 34, 1001-1010, 2000.](#)
- 2182 [Grieshop, A. P., Donahue, N. M., and Robinson, A. L.: Is the gas-particle partitioning](#)  
2183 [in alpha-pinene secondary organic aerosol reversible?, Geophys. Res. Lett., 34,](#)

- 2184 | [L14810, doi:10.1029/2007GL029987, 2007.](#)
- 2185 | [Griffin, R. J., Cocker, D. R., III, Flagan, R. C., and Seinfeld, J. H.: Organic aerosol](#)  
 2186 [formation from the oxidation of biogenic hydrocarbons, J. Geophys. Res., 104, 3555–](#)  
 2187 [3567, 1999.](#)
- 2188 | Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M.,  
 2189 Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann,  
 2190 H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A.,  
 2191 Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prevot, A. S. H., Seinfeld, J.  
 2192 H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact  
 2193 of secondary organic aerosol: current and emerging issues, *Atmospheric Chemistry*  
 2194 *and Physics*, 9, 5155-5236, 2009.
- 2195 | Hamilton, J. F., Lewis, A. C., Bloss, C., Wagner, V., Henderson, A. P., Golding, B. T.,  
 2196 Wirtz, K., Martin-Reviejo, M., and Pilling, M. J.: Measurements of photo-oxidation  
 2197 products from the reaction of a series of alkyl-benzenes with hydroxyl radicals during  
 2198 EXACT using comprehensive gas chromatography, *Atmospheric Chemistry and*  
 2199 *Physics*, 3, 1999-2014, 2003.
- 2200 | Hamilton, J. F., Alfarra, M. R., Wyche, K. P., Ward, M. W., Lewis, A. C., McFiggans, G.  
 2201 B., Good, N., Monks, P. S., Carr, T., White, I. R., and Purvis, R. M.: Investigating the  
 2202 use of secondary organic aerosol as seed particles in simulation chamber  
 2203 experiments, *Atmospheric Chemistry and Physics*, 11, 5917-5929, [10.5194/acp-11-](#)  
 2204 [5917-2011](#), 2011.
- 2205 | Harrison, R. M., Giorio, C., Beddows, D. C. S., and Dall'Osto, M.: Size distribution of  
 2206 airborne particles controls outcome of epidemiological studies, *Science of the Total*  
 2207 *Environment*, 289-293, 2010.
- 2208 | Heal, M. R., Kumar, P., and Harrison, R. M.: Particles, air quality, policy and health,  
 2209 *Chemical Society Reviews*, 6606-6630, [10.1039/c2cs35076a](#), 2012.
- 2210 | [Hennigan, C. J., Miracolo, M. A., Engelhart, G. J., May, A. A., Presto, A. A., Lee, T.,](#)  
 2211 [Sullivan, A. P., McMeeking, G. R., Coe, H., Wold, C. E., Hao, W. -M., Gilman, J. B.,](#)  
 2212 [Kuster, W. C., de Gouw, J., Schichtel, B. A., Collet Jr, J. L., Kreidenweis, S. M., and](#)  
 2213 [Robinson, A. L.: Chemical and physical transformations of organic aerosol from the](#)  
 2214 [photo-oxidation of open biomass burning emissions in an environmental chamber,](#)  
 2215 [Atmos. Chem. Phys., 11, 7669-7686, 2011.](#)
- 2216 | Henze, D. K., and Seinfeld, J. H.: Global Secondary Organic Aerosol from isoprene  
 2217 oxidation, *Geophysical Research Letters*, 33, L09812, 2006.
- 2218 | Heringa, M. F., DeCarlo, P. F., Chirico, R., Tritscher, T., Clairotte, M., Mohr, C., Crippa,  
 2219 M., Slowik, J. G., Pfaffenberger, L., Dommen, J., Weingartner, E., Prevot, A. S. H., and  
 2220 Baltensperger, U.: A new method to discriminate secondary organic aerosols from  
 2221 different sources using high-resolution aerosol mass spectra, *Atmospheric Chemistry*  
 2222 *and Physics*, 12, 2189–2203, doi:10.5194/acp-12-2189-2012, 2012.

2223 [Hildebrandt, L., Donahue, N. M., and Pandis, S. N.: High formation of secondary](#)  
 2224 [organic aerosol from the photo-oxidation of toluene, Atmos. Chem. Phys., 9, 2973–](#)  
 2225 [2986, doi:10.5194/acp-9-2973-2009, 2009.](#)

2226 [Hoffmann, T., Odum, J. R., Bowman, F., Collins, D., Klockow, D., Flagan, R. C., and](#)  
 2227 [Seinfeld, J. H.: Formation of Organic Aerosols from the Oxidation of Biogenic](#)  
 2228 [Hydrocarbons, J. Atmos. Chem., 26, 189-222, 1997.](#)

2229 <http://mcm.leeds.ac.uk/MCM>: in.

2230 IPCC: Intergovernmental Panel on Climate Change, Fourth Assessment Report,  
 2231 Climate Change 2007, 2007.

2232 Jackson, J. E.: PRINCIPAL COMPONENTS AND FACTOR-ANALYSIS .1. PRINCIPAL  
 2233 COMPONENTS, J. Qual. Technol., 12, 201-213, 1980.

2234 Jaoui, M., and Kamens, R. M.: Mass balance of gaseous and particulate products  
 2235 analysis from alpha-pinene/NO<sub>x</sub>/air in the presence of natural sunlight, Journal of  
 2236 Geophysical Research-Atmospheres, 106, 12541-12558, 10.1029/2001jd900005,  
 2237 2001.

2238 [Jaoui, M. and Kamens, R. M.: Gaseous and particulate oxidation products analysis of](#)  
 2239 [a mixture of a  \$\alpha\$ -pinene +  \$\beta\$ -pinene/O<sub>3</sub>/air in the absence of light and  \$\alpha\$ -pinene +  \$\beta\$ -](#)  
 2240 [pinene/NO<sub>x</sub>/air in the presence of natural sunlight, J. Atmos. Chem., 44, 259–297,](#)  
 2241 [2003.](#)

2242 [Jaoui, M., Leungsakul, S., and Kamens, R. M.: Gas and particle products distribution](#)  
 2243 [from the reaction of beta-caryophyllene with ozone, J. Atmos. Chem., 45, 261–287,](#)  
 2244 [2003.](#)

2245 [Jaoui, M., Corse, E., Kleindienst, T. E., Offenburg, J. H., Lewandowski, M., and Edney,](#)  
 2246 [E. O.: Analysis of Secondary Organic Aerosol Compounds from the Photooxidation of](#)  
 2247 [d-Limonene in the Presence of NO<sub>x</sub> and their Detection in Ambient PM<sub>2.5</sub>, Environ.](#)  
 2248 [Sci. Technol., 40\(12\), 3819-3828, doi: 10.1021/es052566z, 2006.](#)

2249 Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of  
 2250 volatile organic compounds: a protocol for mechanism developement, Atmospheric  
 2251 Environment, 3, 81 - 104, 1997.

2252 Jenkin, M. E.: Modelling the formation and composition of secondary organic aerosol  
 2253 from alpha- and beta-pinene ozonolysis using MCM v3, Atmospheric Chemistry and  
 2254 Physics, 4, 1741-1757, 2004.

2255 Jenkin, M. E., Wyche, K. P., Evans, C. J., Carr, T., Monks, P. S., Alfarra, M. R., Barley,  
 2256 M. H., McFiggans, G. B., Young, J. C., and Rickard, A. R.: Development and chamber  
 2257 evaluation of the MCM v3.2 degradation scheme for beta-caryophyllene,  
 2258 Atmospheric Chemistry and Physics, 12, 5275-5308, 10.5194/acp-12-5275-2012,  
 2259 2012.

2260 Kalberer, M., Sax, M., and Samburova, V.: Molecular Size Evolution of Oligomers in  
 2261 Organic Aerosols Collectes in Urban Atmospheres and Generated in a Smog  
 2262 Chamber, *Environmental Science & Technology*, 40, 5917 - 5922, 2006.

2263 Kamens, R. M., and Jaoui, M.: Modelling aerosol formation from  $\alpha$ -pinene + NO<sub>x</sub> in  
 2264 the presence of natural sunlight using gas-phase kinetics and gas-particle  
 2265 partitioning theory, *Environmental Science & Technology*, 35, 1394 - 1405, 2001.

2266 Kanakidou, M., Seinfeld, J. H., Pandis, S. N., Barnes, I., Detener, F. J., Facchini, M. C.,  
 2267 Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C. J., Swietlicki, E., Putaud, J. P.,  
 2268 Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G. K., Winterhalter, R., Myhre, C. L.,  
 2269 Tsigaridis, K., Vignati, E., Stephanou, E. G., and Wilson, J.: Organic aerosol and global  
 2270 climate modelling: a review, *Atmospheric Chemistry and Physics*, 5, 1053-1123,  
 2271 2005.

2272 Khamaganov, V. G., and Hites, R. A.: Rate Constants for the Gas-Phase Reactions of  
 2273 Ozone with Isoprene,  $\alpha$ - and  $\beta$ -Pinene, and Limonene as a Function of Temperature,  
 2274 *The Journal of Physical Chemistry A*, 105, 815-822, DOI: 10.1021/jp002730z, 2001.

2275 Kim, S., Karl, T., Guenther, A., Tyndall, G., Orlando, J., Harley, P., Rasmussen, R., and  
 2276 Apel, E.: Emissions and ambient distributions of Biogenic Volatile Organic  
 2277 Compounds (BVOC) in a ponderosa pine ecosystem: interpretation of PTR-MS mass  
 2278 spectra, *Atmospheric Chemistry and Physics*, 10, 1759-1771, doi:10.5194/acp-10-  
 2279 1759-2010, 2010.

2280 Kjaergaard, H. G., Knap, H. C., Ornsø, K. B., Jørgensen, S., Crounse, J. D., Paulot, F.,  
 2281 and Wennberg, P. O.: Atmospheric Fate of Methacrolein. 2. Formation of Lactone  
 2282 and Implications for Organic Aerosol Production, *Journal of Physical Chemistry A*,  
 2283 116, 5763-5768, [10.1021/jp210853h](https://doi.org/10.1021/jp210853h), 2012.

2284 Kleindienst, T. E., Conner, T. S., McIver, C. D., and Edney, E. O.: [Determination of](#)  
 2285 [secondary organic aerosol products from the photooxidation of toluene and their](#)  
 2286 [implications in ambient PM<sub>2.5</sub>](#), *J. Atmos. Chem.*, 47, 79-100, 2004.

2287 Koch, S., Winterhalter, R., Uherek, E., Koloff, A., Neeb, P., and Moortgat, G. K.:  
 2288 Formation of new particles in the gas phase ozonolysis of monoterpenes,  
 2289 *Atmospheric Environment*, 34, 4031 - 4042, 2000.

2290 Kroll, J. H., Ng, N. L., Murphy, S. M., Flagan, R. C., and Seinfeld, J. H.: Secondary  
 2291 organic aerosol formation from isoprene photooxidation under high-NO<sub>x</sub> conditions,  
 2292 *Geophysical Research Letters*, 32, L18808, 2005.

2293 Kroll, J. H., Ng, N. L., Murphy, S. M., Flagan, R. C., and Seinfeld, J. H.: Secondary  
 2294 organic aerosol formation from isoprene photooxidation, *Environmental Science and*  
 2295 *Technology*, 40, 1869 - 1877, 2006.



2296 Kroll, J. H., and Seinfeld, J. H.: Chemistry of secondary organic aerosol: Formation  
 2297 and evolution of low-volatility organics in the atmosphere, *Atmospheric*  
 2298 *Environment*, 42, 3593-3624, 2008.

2299 Kuppasami, S., Clokie, M. R. J., Panayi, T., Ellis, A. M., and Monks, P. S.: Metabolite  
 2300 profiling of *Clostridium difficile* ribotypes using small molecular weight volatile  
 2301 organic compounds, *Metabolomics*, DOI 10.1007/s11306-014-0692-4, 2014.

2302 [Lanz, V. A., Alfarra, M. R., Baltensperger, U., Buchmann, B., Hueglin, C., and Prevot,](#)  
 2303 [A. S. H.: Source apportionment of submicron organic aerosols at an urban site by](#)  
 2304 [factor analytical modelling of aerosol mass spectra, \*Atmos Chem Phys\*, 7, 1503-1522,](#)  
 2305 [2007.](#)

2306 Larsen, B. R., Di Bella, D., Glasius, M., Winterhalter, R., Jensen, N. R., and Hjorth, J.:  
 2307 Gas-phase OH oxidation of monoterpenes: gaseous and particulate products, *Journal*  
 2308 *of Atmospheric Chemistry*, 38, 231 - 276, 2001.

2309 Lee, A., Goldstein, A. H., Kroll, J. H., Ng, N. L., Varutbangkul, V., Flagan, R. C., and  
 2310 Seinfeld, J. H.: Gas-phase products and secondary aerosol yields from the  
 2311 photooxidation of 16 different terpenes, *Journal of Geophysical Research-*  
 2312 *Atmospheres*, 111, 10.1029/2006jd007050, 2006.

2313 [Li, Y. J., Chen, Q., Guzman, M. I., Chan, C. K., and Martin, S. T.: Second-generation](#)  
 2314 [products contribute substantially to the particle-phase organic material produced by](#)  
 2315 [β-caryophyllene ozonolysis, \*Atmos. Chem. Phys.\*, 11, 121–132, doi:10.5194/acp-11-](#)  
 2316 [121-2011, 2011.](#)

2317 Librando, V., and Tringali, G.: Atmospheric fate of OH initiated oxidation of terpenes.  
 2318 Reaction mechanism of α-pinene degradation and secondary organic aerosol  
 2319 formation, *Journal of Environmental Management*, 75, 275 - 282, 2005.

2320 Limbeck, A., Kulmala, M., and Puxbaum, H.: Secondary organic aerosol formation in  
 2321 the atmosphere via heterogeneous reaction of gaseous isoprene on acid particles,  
 2322 *Geophysical Research Letters*, 30, 2003.

2323 Lin, Y.-H., Zhanga, H., Pye, H. O., Zhanga, Z., Martha, W. J., Parka, S., Arashiroa, M.,  
 2324 Cuia, T., Budisulistiorinia, S. H., Sextona, K. G., Vizuetea, W., Xieb, Y., Lueckenb, D. J.,  
 2325 Pileticb, I. R., Edneyb, E. O., Bartolottic, L. J., Gold, A., and Surratt, S. D.: Epoxide as a  
 2326 precursor to secondary organic aerosol formation from isoprene photooxidation in  
 2327 the presence of nitrogen oxides, *Environmental Science & Technology*, 110, 6718-  
 2328 6723, doi/10.1073/pnas.1221150110, 2013.

2329 Mahalanobis, P. C.: On the generalised distance in statistics, *Proceedings of the*  
 2330 *National Institute of Sciences of India*, 2, 49-55, 1936.

2331 [Marcolli, C., Canagaratna, M. R., Worsnop, D. R., Bahreini, R., de Gouw, J.](#)  
 2332 [A.,Warneke, C., Goldan, P. D., Kuster,W. C.,Williams, E. J., Lerner, B. M., Roberts, J.](#)  
 2333 [M., Meagher, J. F., Fehsenfeld, F. C., Marchewka, M., Bertman, S. B., and](#)

2334 | [Middlebrook, A. M.: Cluster Analysis of the Organic Peaks in Bulk Mass Spectra](#)  
2335 [Obtained During the 2002 New England Air Quality Study with an Aerodyne Aerosol](#)  
2336 [Mass Spectrometer, Atmos. Chem. Phys., 6, 5649–5666, 2006, http://www.atmos-](#)  
2337 [chem-phys.net/6/5649/2006/.](#)

2338 Metzger, A., Dommen, J., Gaeggeler, K., Duplissy, J., Prevot, A. S. H., Kleffmann, J.,  
2339 Elshorbany, Y., Wisthaler, A., and U., B.: Evaluation of 1,3,5-trimethylbenzene  
2340 degradation in the detailed tropospheric chemistry mechanism MCMv3.1 using  
2341 environmental chamber data, Atmospheric Chemistry and Physics, 8, 6453–6468,  
2342 2008.

2343 [Ng, N. L., Kroll, J. H., Keywood, M. D., Bahreini, R., Varutbangkul, V., Flagan, R. C.,](#)  
2344 [Seinfeld, J. H., Lee, A., and Goldstein, A. H.: Contribution of first- versus second-](#)  
2345 [generation products to secondary organic aerosols formed in the oxidation of](#)  
2346 [biogenic hydrocarbons, Environ. Sci. Technol., 40, 2283–2297, 2006.](#)

2347 [Ng, N. L., Kroll, J. H., Chan, A. W. H., Chhabra, P. S., Flagan, R. C., and Seinfeld:](#)  
2348 [Secondary organic aerosol formation from \*m\*-xylene, toluene, and benzene, Atmos.](#)  
2349 [Chem. Phys., 7, 3909–3922, 2007.](#)

2350 [Odum, J. R., Hoffmann, T., Bowman, F., Collins, D., Flagan, R. C., and Seinfeld, J. H.:](#)  
2351 [Gas/particle partitioning and secondary organic aerosol yields, Environ. Sci. Technol.,](#)  
2352 [30, 2580–2585, 1996.](#)

2353 [Paatero, P. and Tapper, U.: Positive Matrix Factorization: a nonnegative factor model](#)  
2354 [with optimal utilization of error estimates of data values, Environmetrics, 5, 111–](#)  
2355 [126, 1994.](#)

2356 [Paatero, P.: Least squares formulation of robust non-negative factor analysis,](#)  
2357 [Chemometr. Intell. Lab., 37, 23–35, 1997.](#)

2359 Paglione, M., Kiendler-Scharr, A., Mensah, A. A., Finessi, E., Giulianelli, L., Sandrini, S.,  
2360 Facchini, M. C., Fuzzi, S., Schlag, P., Piazzalunga, A., Tagliavini, E., Henzing, J. S., and  
2361 S., D.: Identification of humic-like substances (HULIS) in oxygenated organic aerosols  
2362 using NMR and AMS factor analyses and liquid chromatographic techniques,  
2363 Atmospheric Chemistry and Physics, 14, 25–45, doi:10.5194/acp-14-25-2014, 2014.

2364 [Pandis, S. N., Paulson, S. E., Seinfeld, J. H., and Flagan, R. C.: Aerosol Formation in the](#)  
2365 [Photooxidation of Isoprene and Beta-Pinene, Atmos. Environ., 25\(5-6\), 997–1008,](#)  
2366 [doi: 10.1016/0960-1686\(91\)90141-S, 1992.](#)

2367 Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kurten, A., St Clair, J. M., Seinfeld, J. H.,  
2368 and Wennberg, P. O.: Unexpected Epoxide Formation in the Gas-Phase  
2369 Photooxidation of Isoprene, Science, 325, 730–733, [10.1126/science.1172910,](#)  
2370 2009a.

2371 Paulot, F., Crounse, J. D., Kjaergaard, H. G., Kroll, J. H., Seinfeld, J. H., and Wennberg,  
 2372 P. O.: Isoprene photooxidation: new insights into the production of acids and organic  
 2373 nitrates, *Atmospheric Chemistry and Physics*, 9, 1479-1501, 2009b.

2374 Paulot, F., Henze, D. K., and Wennberg, P. O.: Impact of the isoprene photochemical  
 2375 cascade on tropical ozone, *Atmospheric Chemistry and Physics*, 12, 1307-1325,  
 2376 [10.5194/acp-12-1307-2012](https://doi.org/10.5194/acp-12-1307-2012), 2012.

2377 Paulsen, D., Dommen, J., Kalberer, M., Prevot, A. S. H., Richter, R., Sax, M.,  
 2378 Steinbacher, M., Weingartner, E., and Baltensperger, U.: Secondary organic aerosol  
 2379 formation by irradiation of 1,3,5-trimethylbenzene-NO<sub>x</sub>-H<sub>2</sub>O in a new reaction  
 2380 chamber for atmospheric chemistry and physics, *Environmental Science &*  
 2381 *Technology*, 39, 2668-2678, 2005.

2382 [Presto, A. A., HuffHartz, K. E., and Donahue, N. M.: Secondary Organic Aerosol](#)  
 2383 [Production from Terpene Ozonolysis. 2. Effect of NO<sub>x</sub> Concentration, \*Environ. Sci.\*](#)  
 2384 [Technol.](#), 39, 7046–7054, 2005.

2385 Reinnig, M. C., Muller, L., Warnke, J., and Hoffmann, T.: Characterization of selected  
 2386 organic compound classes in secondary organic aerosol from biogenic VOCs by  
 2387 HPLC/MS, *Analytical Bioanalytical Chemistry*, 391, 171 - 182, 2008.

2388 [Rohrer, F., Bohn, B., Brauers, T., Bruning, D., Johnen, F. -J., Wahner, A., and](#)  
 2389 [Kleffmann, J.: Characterisation of the photolytic HONO-source in the atmosphere](#)  
 2390 [simulation chamber SAPHIR, \*Atmos. Chem. Phys.\*, 5, 2189-2201, 2005.](#)

2391 Rickard, A. R., Wyche, K. P., Metzger, A., Monks, P. S., Ellis, A. M., Dommen, J.,  
 2392 Baltensperger, U., Jenkin, M. E., and Pilling, M. J.: Gas phase precursors to  
 2393 anthropogenic secondary organic aerosol Using the Master Chemical Mechanism to  
 2394 probe detailed observations of 1,3,5-trimethylbenzene photo-oxidation,  
 2395 *Atmospheric Environment*, 44, 5423-5433, [10.1016/j.atmosenv.2009.09.043](https://doi.org/10.1016/j.atmosenv.2009.09.043), 2010.

2396 Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the  
 2397 developement of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric  
 2398 degradation of non-aromatic volatile organic compounds, *Atmospheric Chemistry*  
 2399 *and Physics*, 3, 161 - 180, 2003.

2400 Shu, Y. H., Kwok, E. S. C., Tuazon, E. C., Atkinson, R., and Arey, J.: Products of the gas-  
 2401 phase reactions of linalool with OH radicals, NO<sub>3</sub> radicals, and O-3, *Environmental*  
 2402 *Science & Technology*, 31, 896-904, [10.1021/es960651o](https://doi.org/10.1021/es960651o), 1997.

2403 Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T.,  
 2404 Muller, J. F., Kuhn, U., Stefani, P., and Knorr, W.: Global data set of biogenic VOC  
 2405 emissions calculated by the MEGAN model over the last 30 years, *Atmospheric*  
 2406 *Chemistry and Physics*, 14, 9317-9341, [10.5194/acp-14-9317-2014](https://doi.org/10.5194/acp-14-9317-2014), 2014.

2407 [Slowik, J. G., Vlasenko, A., McGuire, M., Evans, G. J., and Abbatt, J. P. D.:  
2408 Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-  
2409 MS measurements at an urban site, \*Atmos. Chem Phys.\*, 10, 1969-1988, 2010](#)

2410 Sousa, C., Novais, A., Magalhaes, A., Lopes, J., and Peixe, L.: Diverse high-risk B2 and  
2411 D *Escherichia coli* clones depicted by Fourier Transform Infrared Spectroscopy, *Sci*  
2412 *Rep*, 3, 8, [327810.1038/srep03278](#), 2013.

2413 Sun, X., Zhang, C., Zhao, Y., Bai, J., and Hea, M.: Kinetic study on the linalool  
2414 ozonolysis reaction in the atmosphere, *Canadian Journal of Chemistry*, 90, doi:  
2415 10.1139/v2012-001, 2012.

2416 Surratt, J. D., Murphy, S. M., Kroll, J. H., Ng, N. L., Hilderbrandt, L., Sorooshian, A.,  
2417 Szmigielski, R., Vermeylen, R., Maenhaut, W., Claeys, M., Flagen, R., and Seinfeld, J.  
2418 H.: Chemical Composition of Secondary Organic Aerosol Formed from the  
2419 Photooxidation of Isoprene, *Journal of Physical Chemistry A*, 110, 9665-9690, 2006.

2420 Surratt, J. D., Chan, A. W. H., Eddingsaas, N. C., Chan, M. N., Loza, C. L., Kwan, A. J.,  
2421 Hersey, S. P., Flagan, R. C., Wennberg, P. O., and Seinfeld, J. H.: Reactive  
2422 intermediates revealed in secondary organic aerosol formation from isoprene, *Proc.*  
2423 *Natl. Acad. Sci. U. S. A.*, 107, 6640-6645, [10.1073/pnas.0911114107](#), 2010.

2424 Surratt, J. D.: Radical regeneration from isoprene, *Nat. Geosci.*, 6, 995-996, 2013.

2425 Tolocka, M. P., Jang, M., Ginter, J. M., Cox, F. J., Kamens, R. M., and Johnston, M. V.:  
2426 Formation of Oligomers in Secondary Organic Aerosol, *Environmental Science &*  
2427 *Technology*, 38, 1428-1434, [10.1021/es035030r](#), 2004.

2428 [Wagner, V., Jenkin, M. E., Saunders, S. M., Stanton, J., Wirtz, K., and Pilling, M. J.:  
2429 Modelling of the photooxidation of toluene: conceptual validating detailed  
2430 mechanisms, \*Atmos. Chem. Phys.\*, 3, 89-106, 2003.](#)

2431 Wang, J. L., Chew, C., Chang, C. Y., Liao, W. C., Lung, S. C. C., Chen, W. N., Lee, P. J.,  
2432 Lin, P. H., and Chang, C. C.: Biogenic isoprene in subtropical urban settings and  
2433 implications for air quality, *Atmospheric Environment*, 79, 369-379,  
2434 [10.1016/j.atmosenv.2013.06.055](#), 2013.

2435 [Winterhalter, R., Herrmann, F., Kanawati, B., Nguyen, T. L., Peeters, J., Vereecken, L.,  
2436 and Moortgat, G. K.: The gas-phase ozonolysis of  \$\beta\$ -caryophyllene \(C<sub>15</sub>H<sub>24</sub>\). Part I:  
2437 an experimental study, \*Phys. Chem. Chem. Phys.\*, 11, 4152-4172, 2009.](#)

2438 Worton, D. R., Surratt, J. D., LaFranchi, B. W., Chan, A. W. H., Zhao, Y. L., Weber, R. J.,  
2439 Park, J. H., Gilman, J. B., de Gouw, J., Park, C., Schade, G., Beaver, M., St Clair, J. M.,  
2440 Crounse, J., Wennberg, P., Wolfe, G. M., Harrold, S., Thornton, J. A., Farmer, D. K.,  
2441 Docherty, K. S., Cubison, M. J., Jimenez, J. L., Frossard, A. A., Russell, L. M.,  
2442 Kristensen, K., Glasius, M., Mao, J. Q., Ren, X. R., Brune, W., Browne, E. C., Pusede, S.  
2443 E., Cohen, R. C., Seinfeld, J. H., and Goldsteint, A. H.: Observational Insights into

2444 Aerosol Formation from Isoprene, *Environmental Science & Technology*, 47, 11403-  
 2445 11413, [10.1021/es4011064](https://doi.org/10.1021/es4011064), 2013.

2446 Wyche, K. P., Blake, R. S., Ellis, A. M., Monks, P. S., Brauers, T., Koppmann, R., and  
 2447 Apel, E. C.: Technical note: Performance of Chemical Ionization Reaction Time-of-  
 2448 Flight Mass Spectrometry (CIR-TOF-MS) for the measurement of atmospherically  
 2449 significant oxygenated volatile organic compounds, *Atmospheric Chemistry and  
 2450 Physics*, 7, 609-620, 2007.

2451 Wyche, K. P., Monks, P. S., Ellis, A. M., Cordell, R. L., Parker, A. E., Whyte, C.,  
 2452 Metzger, A., Dommen, J., Duplissy, J., Prevot, A. S. H., Baltensperger, U., Rickard, A.  
 2453 R., and Wulfert, F.: Gas phase precursors to anthropogenic secondary organic  
 2454 aerosol: detailed observations of 1,3,5-trimethylbenzene photooxidation,  
 2455 *Atmospheric Chemistry and Physics*, 9, 635-665, 2009.

2456 Wyche, K. P., Ryan, A. C., Hewitt, C. N., Alfarra, M. R., McFiggans, G. B., Carr, T.,  
 2457 Monks, P. S., Smallbone, K. L., Capes, G., Hamilton, J. F., Pugh, T. A. M., and  
 2458 MacKenzie, A. R.: Emissions of biogenic volatile organic compounds and subsequent  
 2459 photochemical production of secondary organic aerosol in mesocosm studies of  
 2460 temperate and tropical plant species, *Atmospheric Chemistry and Physics*, In Press,  
 2461 2014.

2462 Yu, J. Z., Cocker, D. R., Griffin, R. J., Flagan, R. C., and Seinfeld, J. H.: Gas-phase ozone  
 2463 oxidation of monoterpenes: Gaseous and particulate products, *Journal of  
 2464 Atmospheric Chemistry*, 34, 207-258, [10.1023/a:1006254930583](https://doi.org/10.1023/a:1006254930583), 1999.

2465 Yu, Y., Ezell, M. J., Zelenyuk, A., Imre, D., Alexander, L., Ortega, J., D'Anna, B.,  
 2466 Harmon, C. W., Johnson, S. N., and Finlayson-Pitts, B. J.: Photooxidation of alpha-  
 2467 pinene at high relative humidity in the presence of increasing concentrations of NO<sub>x</sub>,  
 2468 *Atmospheric Environment*, 42, 5044-5060, [10.1016/j.atmosenv.2008.02.026](https://doi.org/10.1016/j.atmosenv.2008.02.026), 2008.

2469 Zador, J., Turanyi, T., Wirtz, K., and Pilling, M. J.: Measurement and investigation of  
 2470 chamber radical sources in the EUropean PHoto REactor (EUPHORE), *Journal of  
 2471 Atmospheric Chemistry*, 55, 147 - 166, 2006.

2472 [Zhang, Q., Alfarra, M. R., Worsnop, D. R., Allan, J. D., Coe, H., Canagaratna, M. R., and](#)  
 2473 [Jimenez, J. L.: Deconvolution and Quantification of Hydrocarbon-like and](#)  
 2474 [Oxygenated Organic Aerosols Based on Aerosol Mass Spectrometry, \*Environmental\*](#)  
 2475 [Science & Technology](#), 39, 4938-4952, 2005.

2476 [Zhang, Q., Jimenez, J. L., Canagaratna, M. R., Allan, J. D., Coe, H., Ulbrich, I., Alfarra,](#)  
 2477 [M. R., Takami, A., Middlebrook, A. M., Sun, Y. L., Dzepina, K., Dunlea, E., Docherty, K.,](#)  
 2478 [DeCarlo, P. F., Salcedo, D., Onasch, T., Jayne, J. T., Miyoshi, T., Shimono, A.,](#)  
 2479 [Hatakeyama, S., Takegawa, N., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S.,](#)  
 2480 [Weimer, S., Demerjian, K., Williams, P., Bower, K., Bahreini, R., Cottrell, L., Griffin, R.](#)  
 2481 [J., Rautiainen, J., Sun, J. Y., Zhang, Y. M., and Worsnop, D. R.: Ubiquity and](#)  
 2482 [dominance of oxygenated species in organic aerosols in anthropogenically-](#)  
 2483

2484 | [influenced Northern Hemisphere midlatitudes, Geophysical Research Letters, 34,](#)  
2485 | [L13801, doi:10.1029/2007GL029979, 2007.](#)


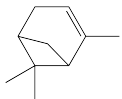
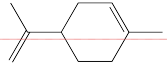
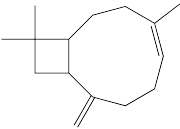
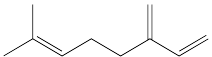
2486 |

2487 |

Kevin Wyche 6/5/2015 14:54  
**Formatted:** Indent: First line: 0 cm,  
Space Before: Auto, After: 12 pt

10. Tables

Table 1: Summary of experiments conducted

Experiment ID	Precursor	Structure	$k(\text{OH}) / k(\text{O}_3)^4 / \text{cm}^3 \text{ molec}^{-1} \text{ s}^{-1}$	Experiment Type (no.)	VOC/ $\text{NO}_x$ Range	RH / % Range
ISOP1 – 10 <sup><u>M</u>,<u>P</u></sup>	isoprene		$9.9 \times 10^{-11} / 1.2 \times 10^{-17}$	Photooxidation (10)	1.3 – 20.0	49 – 72
APIN1 – 4 <sup><u>2</u>,<u>3</u>,<u>M</u>,<u>P</u></sup>	$\alpha$ -pinene		$5.3 \times 10^{-11} / 8.4 \times 10^{-17}$	Photooxidation (4)	1.3 – 2.0 <sup>1</sup>	49 – 73
LIM1 – 6 <sup><u>2</u>,<u>3</u>,<u>M</u>,<u>E</u>,<u>P</u></sup>	limonene		$1.7 \times 10^{-10} / 2.1 \times 10^{-16}$	Photooxidation (6)	1.4 – 2.0 <sup>1</sup>	50 <sup>1</sup> – 82
BCARY1 – 10 <sup><u>2</u>,<u>3</u>,<u>M</u></sup>	$\beta$ -caryophyllene		$2.0 \times 10^{-10} / 1.2 \times 10^{-14}$	Photooxidation (10)	0.6 – 2.0 <sup>1</sup>	50 <sup>1</sup> – 72
MYRC1 – 2 <sup><u>2</u>,<u>3</u>,<u>M</u></sup>	myrcene		$2.1 \times 10^{-10} / 4.7 \times 10^{-16}$	Photooxidation (2)	1.4 – 1.9	52 – 54

Kevin Wyche 16/4/2015 11:31

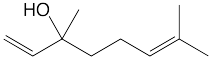
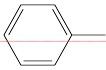
Formatted Table

Kevin Wyche 16/4/2015 11:30

Formatted: Superscript

Kevin Wyche 16/4/2015 11:31

Formatted: Superscript

Experiment ID	Precursor	Structure	$k(\text{OH}) / k(\text{O}_3)^4 / \text{cm}^3 \text{ molec}^{-1} \text{ s}^{-1}$	Experiment Type (no.)	VOC/NO <sub>x</sub> Range	RH / % Range
LINA1 – 2 <sup>2,3,M</sup>	linalool		$1.6 \times 10^{-10} / 4.5 \times 10^{-16}$	Photooxidation (2)	1.4 – 2.6	42 – 47
BIR1 – 2 <sup>M</sup>	birch trees	Multiple emissions <sup>5</sup>	Multiple emissions	Mesocosm Photooxidation (2)	5.5 – 5.6	73 – 84
FIG1 – 2 <sup>M</sup>	fig trees	Multiple emissions <sup>5</sup>	Multiple emissions	Mesocosm Photooxidation (2)	2.7 – 9.4	65 – 75
TOL1 – 5 <sup>E</sup>	toluene		$3.7 \times 10^{-12} / -$	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

Kevin Wyche 16/4/2015 11:31

Formatted Table

Kevin Wyche 16/4/2015 11:31

Formatted: Superscript

Kevin Wyche 16/4/2015 11:32

Formatted: Superscript

Kevin Wyche 16/4/2015 11:32

Formatted: Superscript



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 <sup>3</sup> , 3(H) x 3(L) x 2(W) m	1 x 6 kW Xe arc lamp Bank of halogen lamps	$\lambda$ range = 290 – 800 nm $j_{NO_2} = 6 \times 10^{-4} \text{ s}^{-1}$ (290 – 422 nm)
EUPHORE	FEP Teflon	Outdoor	200 m <sup>3</sup> , (hemispherical)	Solar	Solar; 75 % transmission at 290 nm, 85 % transmission > 320 nm $j_{NO_2} = \sim 5 - 9 \times 10^{-3} \text{ s}^{-1}$
PSISC	FEP DuPont Tedlar	Indoor	27 m <sup>3</sup> , 3(H) x 3(L) x 3(W) m	4 x 4 kW Xenon arc lamps	$\lambda$ range = 290 – 800 nm $j_{NO_2} = 0.12 \text{ min}^{-1}$

**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.

Kevin Wyche 16/4/2015 15:39

Formatted: Font:Bold

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

	<u>limononaldehyde</u>	<u>limonene</u>				
<u>139</u>	<u>limonaketone</u>	<u>limonene</u>	=	=	=	=
<u>151</u>	<u>piononaldehyde / limononaldehyde</u>	<u>α-pinene / limonene</u>	=	=	=	=
<u>169</u>	<u>piononaldehyde / limononaldehyde</u>	<u>α-pinene / limonene</u>	<u>pinalic-3-acid / ketolimonon- aldehyde + limonic acid</u>	<u>α-pinene / limonene</u>	=	=
<u>183</u>	=	=	<u>pinonic acid / limonic acid + 7- hydroxylimonon- aldehyde</u>	<u>α-pinene / limonene</u>	=	=
<u>185</u>	=	=	<u>pinic acid / limonic acid</u>	<u>α-pinene / limonene</u>	=	=
<u>199</u>	=	=	<u>C<sub>9</sub>H<sub>11</sub>O<sub>5</sub></u>	<u>myrcene</u>	=	=
<u>215</u>	=	=	<u>C<sub>10</sub>H<sub>15</sub>O<sub>5</sub></u>	<u>myrcene</u>	=	=
<u>235</u>	<u>β-caryophyllene secondary ozonide + isomers</u>	<u>β-caryophyllene</u>	=	=	=	=
<u>227</u>	=	=	<u>C<sub>10</sub>H<sub>11</sub>O<sub>6</sub></u>	<u>myrcene</u>	=	=
<u>237</u>	<u>β-caryophyllene aldehyde</u>	<u>β-caryophyllene</u>	<u>3-[2,2-dimethyl-4- (1-methylene-4-oxo- butyl)-cyclobutyl]-</u>	<u>β-caryophyllene</u>	=	=

			<u>propanoic acid</u>			
<u>251</u>	=	=	<u>β-caryophyllonic acid</u>	<u>β-caryophyllene</u>	=	=
<u>253</u>	<u>β-caryophyllene secondary ozonide + isomers</u>	<u>β-caryophyllene</u>	=	=	=	=
<u>255</u>	=	=	<u>4-(2-(2-carboxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid</u>	<u>β-caryophyllene</u>	=	=
<u>267</u>	=	=	<u>β-14-hydroxy-caryophyllonic acid + β-10-hydroxy-caryophyllonic acid</u>	<u>β-caryophyllene</u>	=	=
<u>271</u>	=	=	<u>4-(2-(3-hydroperoxy-3-oxopropyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid / 4-(2-(2-carboxy-1-hydroxyethyl)-3,3-dimethylcyclobutyl)-4-oxobutanoic acid</u>	<u>β-caryophyllene</u>	=	=
<u>321</u>	=	=	<u>[M-H<sub>2</sub>+FA+Na]<sup>+</sup> C<sub>12</sub>H<sub>14</sub>O<sub>6</sub> / C<sub>13</sub>H<sub>18</sub>O<sub>5</sub></u>	<u>myrcene</u>	=	=

**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

Kevin Wyche 16/4/2015 15:39

Deleted: 3

## 11. Figures

**Figure 1:** (a) NO<sub>x</sub>, O<sub>3</sub>, 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;  $\alpha$ -pinene and limonene), sesquiterpene ( $\beta$ -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<sub>4</sub>-hydroxycarbonyls/methacrylic acid,  $m/z$  237 =  $\beta$ -caryophyllene aldehyde, 235 and 253 =  $\beta$ -caryophyllene secondary ozonide (and isomers thereof),  $m/z$  107, 151 and 169 = pionaldehyde 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.

Kevin Wyche 16/4/2015 12:34

Formatted: Font:Not Bold

**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 ( $\alpha$ -pinene and limonene), sesquiterpene ( $\beta$ -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 ( $\alpha$ -pinene and limonene), sesquiterpene ( $\beta$ -

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 ( $\alpha$ -pinene and limonene), yellow = birch, light blue = straight chain monoterpenes (myrcene and linalool) and dark blue = sesquiterpene ( $\beta$ -caryophyllene).

**Figure 6:** scores plot of the first three latent variables derived from the PLS-DA model analysis of the isoprene, cyclic monoterpene ( $\alpha$ -pinene and limonene), sesquiterpene ( $\beta$ -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 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,  $\alpha$ -pinene,  $\beta$ -caryophyllene and



myrcene, and the associated mass transfer to the particle-phase. Red arrows and text = “high” NO<sub>x</sub> pathways, green arrows and text = “low NO<sub>x</sub>” 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<sub>2</sub>, NO, HO<sub>2</sub> and/or RO<sub>2</sub>), leading to structurally similar products containing different functional groups. *α-pinene mechanism* – X = OH, =O, OOH or ONO<sub>2</sub>; Y = CHO or C(O)OH; Z = OH, OOH or ONO<sub>2</sub>. *β-caryophyllene mechanism* – X = CH<sub>2</sub>OH(OH), CH<sub>2</sub>OH(OOH), CH<sub>2</sub>OH(ONO<sub>2</sub>) or =O. *Myrcene mechanism* – Y = OOH or ONO<sub>2</sub>; 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<sub>x</sub>, high VOC/NO<sub>x</sub> ratio (red diamonds), moderate NO<sub>x</sub>, medium VOC/NO<sub>x</sub> ratio (green squares) and high NO<sub>x</sub>, low VOC/NO<sub>x</sub> 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.

Kevin Wyche 7/5/2015 14:59

Deleted: 10

Kevin Wyche 15/4/2015 16:36

Deleted: /