Supplement of

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

K. P. Wyche et al.

Correspondence to: K. P. Wyche (k.p.wyche@brighton.ac.uk)

S1. Potential positioning of aromatic terpenes in the statistical space

A small fraction of biogenic terpenes are aromatic in nature, possessing a 6-member benzene ring structure. Such species are typically emitted by vegetation as a response to attack or stress (*e.g.* Niinemets, Trends Plant Sci., 15, 145–153, 2010).

Using the MEGAN2.1 model, Guenther *et al.*, (2012; DOI: 10.5194/gmd-5-1471-2012) estimated global annual emissions of ~ 535 Tg yr^{-1} for isoprene and a sum of ~ 147 Tg yr^{-1} for non-aromatic terpenes; total sesquiterpene emissions were of the order 29 Tg yr^{-1}. The authors reported aromatic terpene emissions grouped with ~ thirty "other" monoterpenes. The results implied that total annual emissions of aromatic monoterpenes would be < 15 Tg yr^{-1}, *i.e.* only 2 % of Σ(isoprene, major non-aromatic terpenes and sesquiterpenes). Estimated annual toluene emissions are also not reported individually by Guenther *et al.*, but are assumed to be relatively minor, being reported with an "additional 11 stress VOC(s)", with a total annual emission value of 7.8 Tg yr^{-1}. In later work, Sindalevora *et al.* (2014; DOI: 10.5194/acp-14-9317-2014) use the MEGAN21-MERRA model ensemble to report mean isoprene, Σ(monoterpenes), Σ(sesquiterpenes) and toluene annual global emissions to be of the order 594, 95, 20 and 1.5 Tg yr^{-1}, respectively, *i.e.* toluene emissions were estimated to be only ~ 0.2 % of Σ(isoprene, major non-aromatic terpenes and sesquiterpenes) emissions.

In summary, biogenic aromatic compounds appear to comprise only a small fraction of total average annual terpene emissions and it is therefore likely under most conditions that their oxidation products would be present at such low abundance compared to those discussed within our work, that they would not cause severe spectral/statistical interference.

Under exceptional circumstances where aromatic terpenes are present in significant abundance, it is possible that their spectra could group in the vicinity of those of other monoterpenes (rather than predominantly anthropogenic VOCs) in statistical space, on account of them possessing spectra with some similar features, *e.g.* Lee *et al.* (2006; DOI: 10.1029/2006JD007050) investigated the PTR-MS spectra of methyl chavicol and reported the presence of a parent ion of *m/z* 149 and major product

ions of *m/z* 151, 137, 121 and 109. Of these ions *m/z* 149, 151 and 137 do not appear in toluene proton transfer reaction mass spectra and moreover, the ions of *m/z* 151, 137 and 109 are also significant contributors to monoterpene PTR-MS oxidation spectra.

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

Results: α-terpinene oxidation produces a significant amount of α-terpinaldehyde, which like pinonaldehyde (from α-pinene photooxidation) possesses a spectral fingerprint including ions of *m/z* 169, 151, 123 and 107. As with other non-aromatic terpenes, α-terpinene photooxidation PTR-MS spectra also contains ions of *m/z* 139 (*c.f.* limonaketone from limonene) and 171 (also observed in α-pinene phtooxidation spectra), as well as more unique features, such as an *m/z* 143 ion of significant abundance (also see Lee *et al.,* 2006, DOI: 10.1029/2006JD007050). In combination, such features help to group the α-terpinene photooxidation spectrum in the vicinity of other terpenes, separate in statistical space from the aromatic toluene photooxidation spectra (see Figure S1).
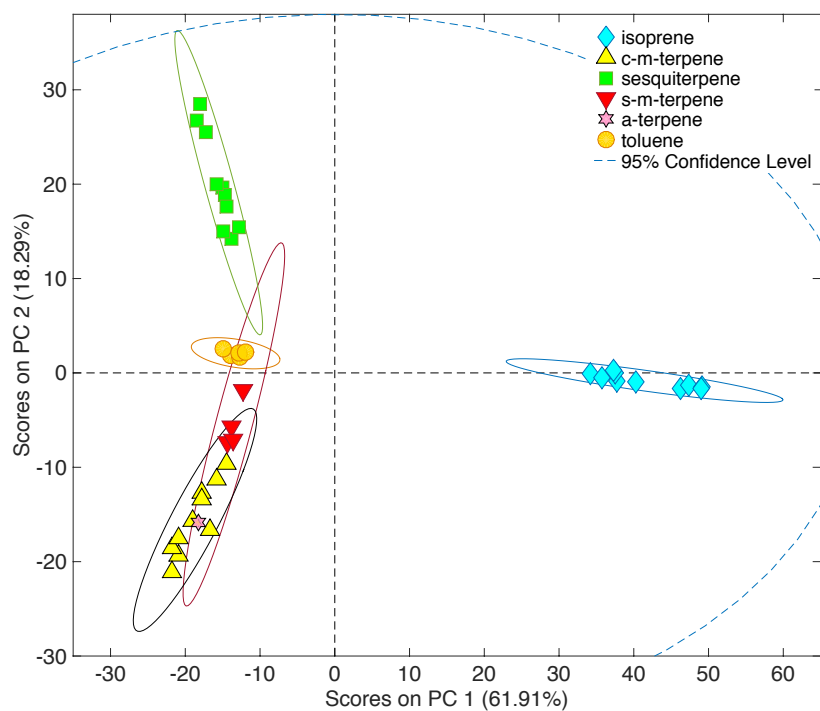
Figure S1: PCA scores plot of the second vs. first principal components derived from the PCA analysis of the toluene test set using the PCA model developed from the isoprene, cyclic monoterpene (α-pinene and limonene), sesquiterpene (β-caryophyllene), straight chain monoterpene (myrcene and linalool) and α-terpinene chamber data. Classification confidence levels = 95 %.