

Manuscript prepared for Atmos. Chem. Phys. Discuss.
with version 1.3 of the L^AT_EX class copernicus_discussions.cls.
Date: 24 November 2011

Supplementary material for "The sensitivity of secondary organic aerosol (SOA) component partitioning to the predictions of component properties- part 3: investigation of condensed compounds generated by a near-explicit model of VOC oxidation"

M. H. Barley, D. Topping, D. Lowe, S. Utembe, and G. McFiggans

Centre for Atmospheric Sciences, School of Earth Atmospheric & Environmental Sciences,
University of Manchester, Manchester M13 9PL UK

Correspondence to: G. McFiggans (g.mcfiggans@manchester.ac.uk)

1 Introduction

This supplementary material is in three parts:-

- Section 2:- Extra figures referred to in the main paper
- Section 3:- The use of SMARTS to provide groups from molecular structures.
- Section 4:- The correction to the Nannoolal Tb method for dicarboxylic acids.

2 Extra figures referred to in the main paper

Figures S1 and S2 show predicted particulate properties across a range of emissions with the BVOC level held constant at 0.01 times that of the standard scenario (S1), and with the AVOC level held constant at the same value (S2).

Figure S5 shows the ratio of SOA Mass(Non-Ideal) to SOA Mass(Ideal) (inversely related to the average effective activity coefficient) against %RH for the scenarios used in Fig. 2

Figure S6 shows very similar scatter for two extreme conditions for the standard scenario and they both look very similar to Figure 3 in the main paper. This supports the change in compound ordering being relatively insensitive to the conditions used for the partitioning.

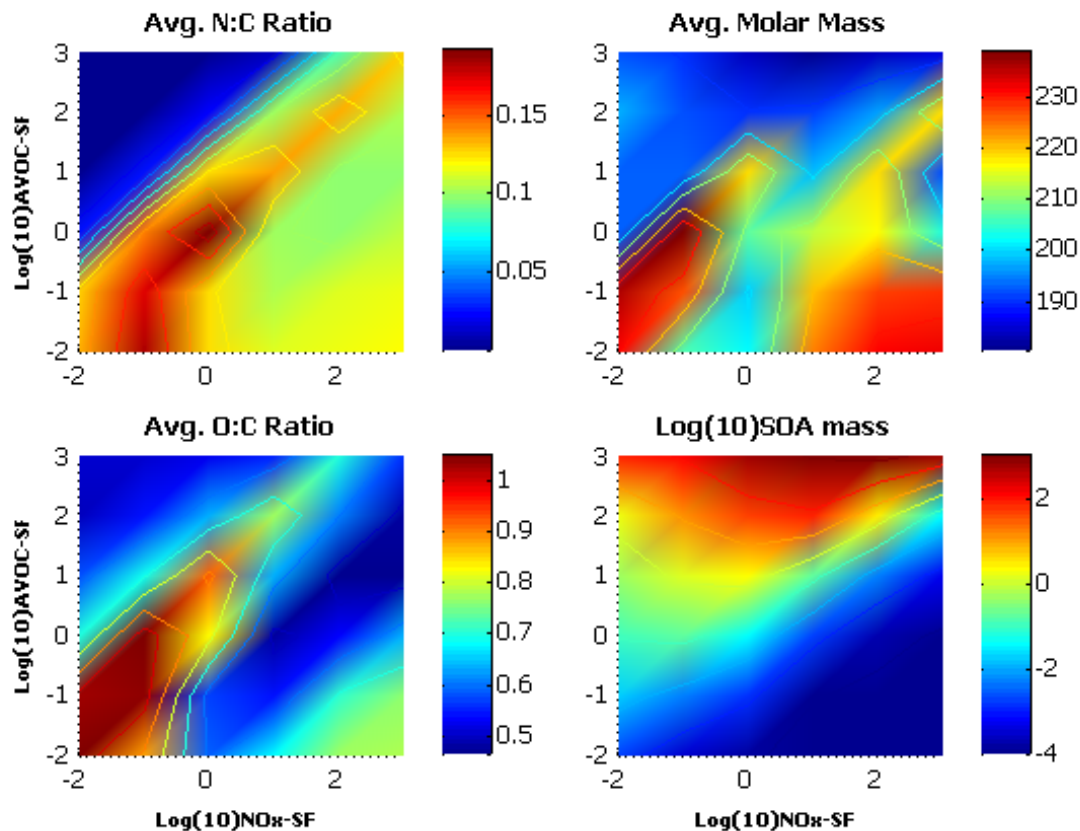


Fig. S1. Surface plots of key properties for those scenarios with very low biogenic inputs (BVOC = 0.01x typical UK emissions) at conditions $T = 293.15\text{K}$, $\%RH = 70$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core, p^0 by N/N-VP, $\gamma_i = 1$. SOA mass is in $\mu\text{grams.m}^{-3}$)

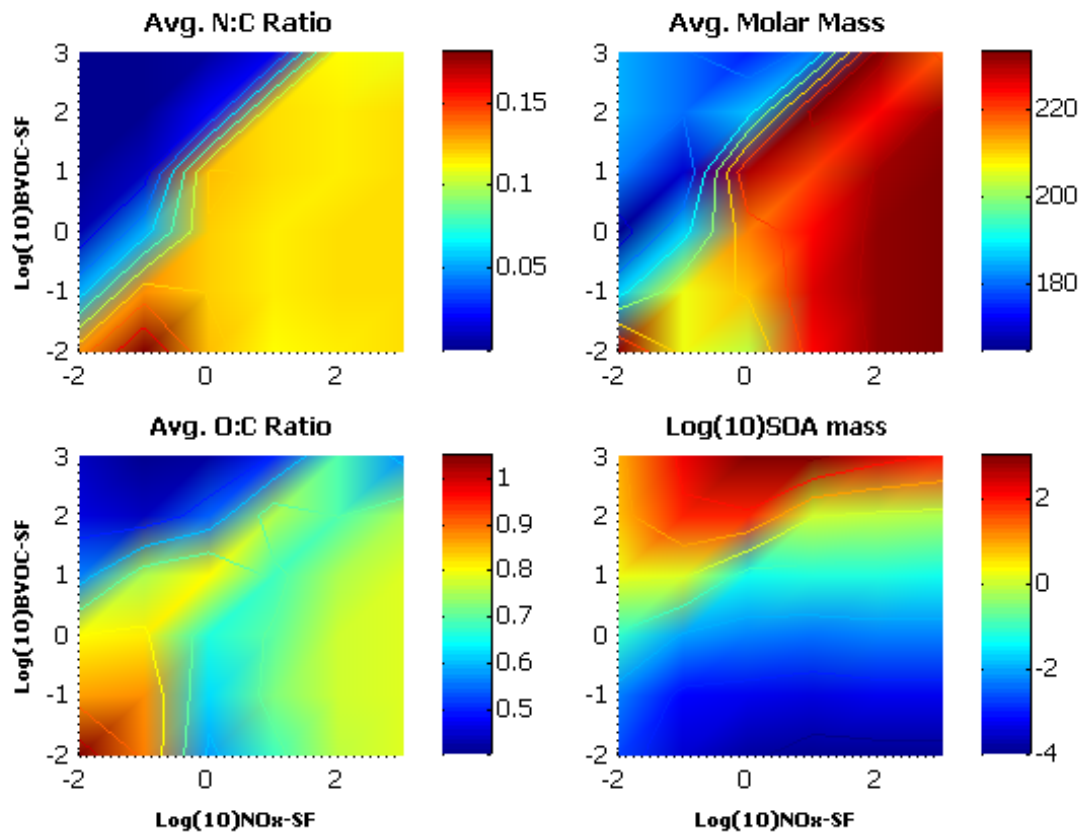


Fig. S2. Surface plots of key properties for those scenarios with very low anthropogenic inputs (AVOC = 0.01x typical UK emissions) at conditions $T = 293.15\text{K}$, $\%RH = 70$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core, p^0 by N/N-VP, $\gamma_i = 1$. SOA mass is in $\mu\text{grams.m}^{-3}$

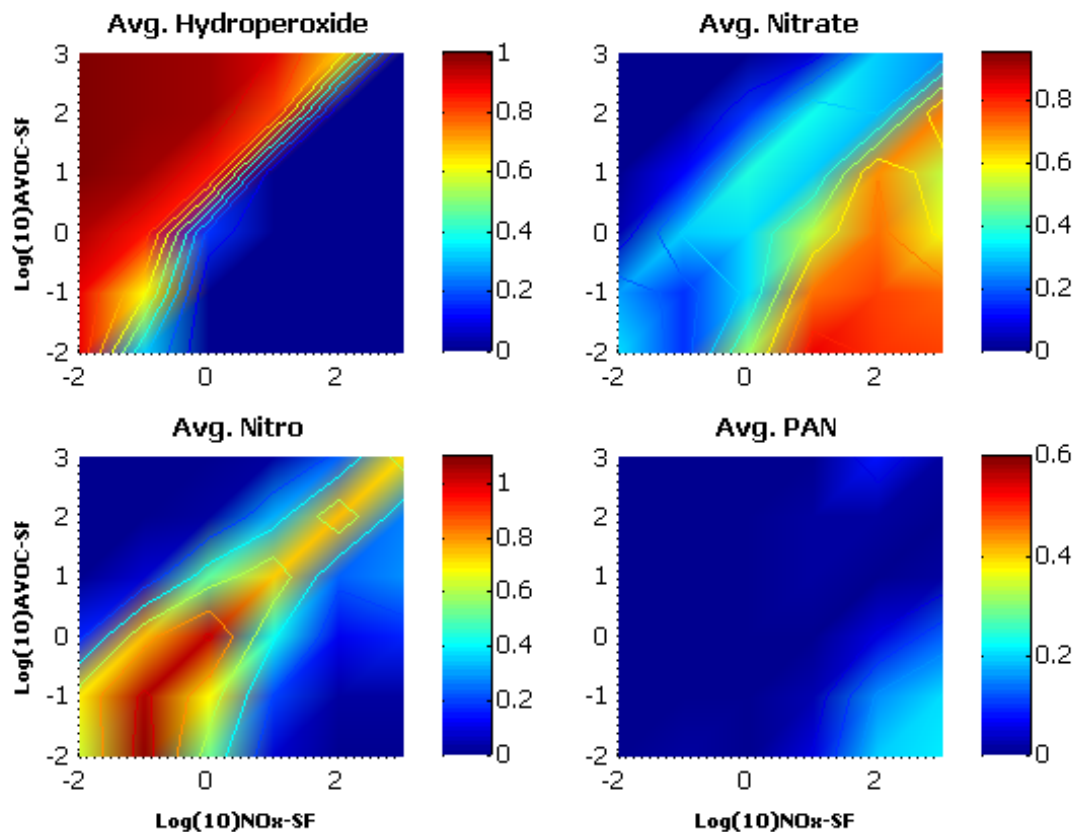


Fig. S3. Surface plots of the distribution of key functional groups for those scenarios with very low biogenic inputs (BVOC = 0.01x typical UK emissions) at conditions $T = 293.15\text{K}$, $\%RH = 70$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core, p^0 by N/N-VP, $\gamma_i = 1$. The plots show the average number of the specified functional group per molecule in the SOA

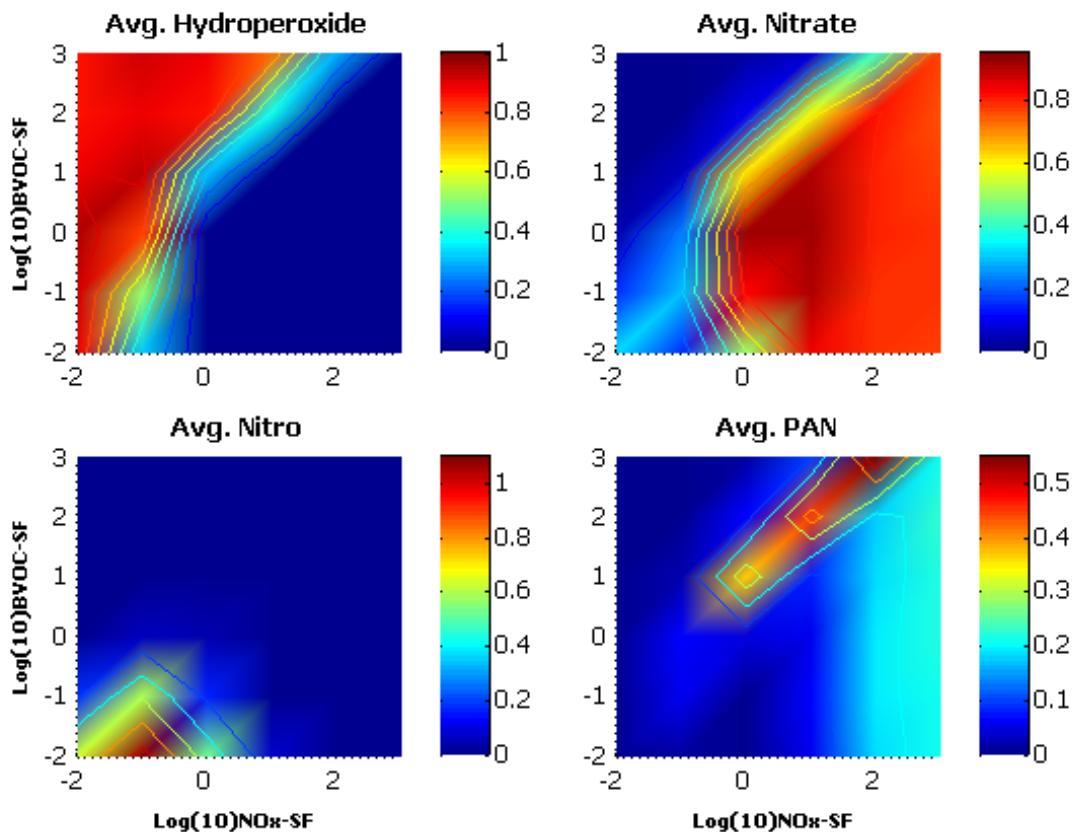


Fig. S4. Surface plots of the distribution of key functional groups for those scenarios with very low anthropogenic inputs ($\text{AVOC} = 0.01 \times$ typical UK emissions) at conditions $T = 293.15\text{K}$, $\%RH = 70$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core, p^0 by N/N-VP, $\gamma_i = 1$. The plots show the average number of the specified functional group per molecule in the SOA

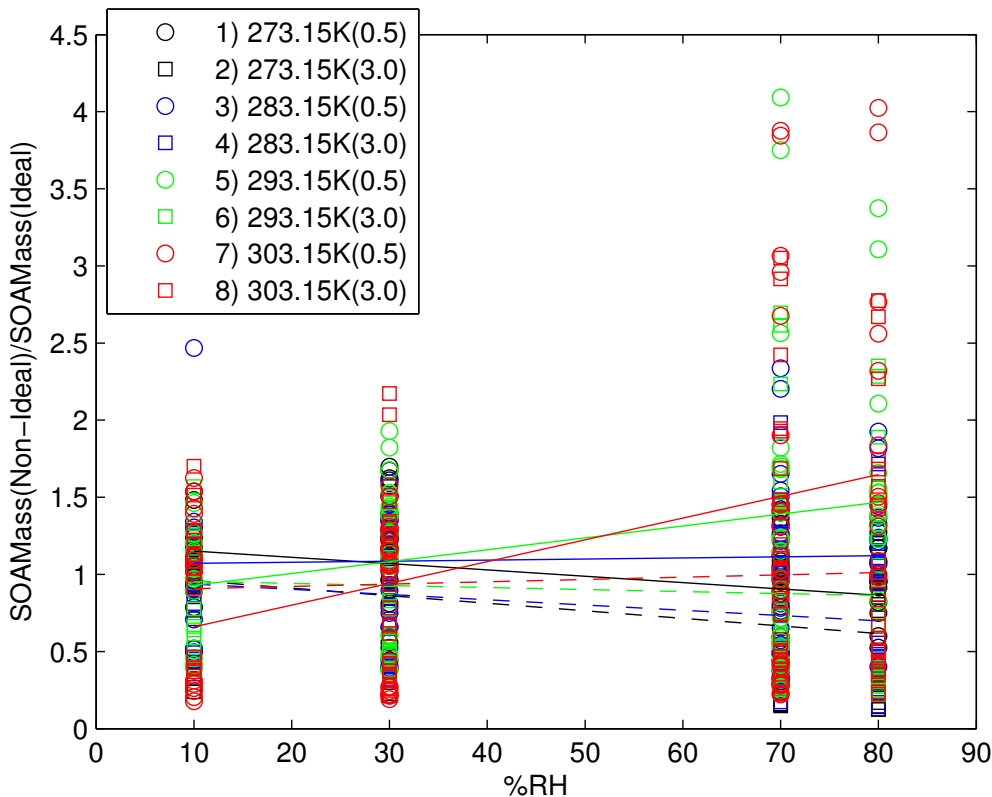


Fig. S5. Plot of the ratio of SOA Mass(Non-Ideal) to SOA Mass(Ideal) (inversely related to the average effective activity coefficient) against %RH for the scenarios used in Fig. 2 (p^0 by N/N-VP); with the data classified into 8 sets by temperature and core size (in brackets:- units of $\mu\text{grams.m}^{-3}$). Individual trend lines for each set are shown (solid lines core = $0.5\mu\text{grams.m}^{-3}$; dashed lines core = $3.0\mu\text{grams.m}^{-3}$ and the R^2 values are:- 1) 0.10261, 2) 0.12644, 3) 0.00135, 4) 0.03958, 5) 0.07005, 6) 0.00406, 7) 0.14892, 8) 0.00412.

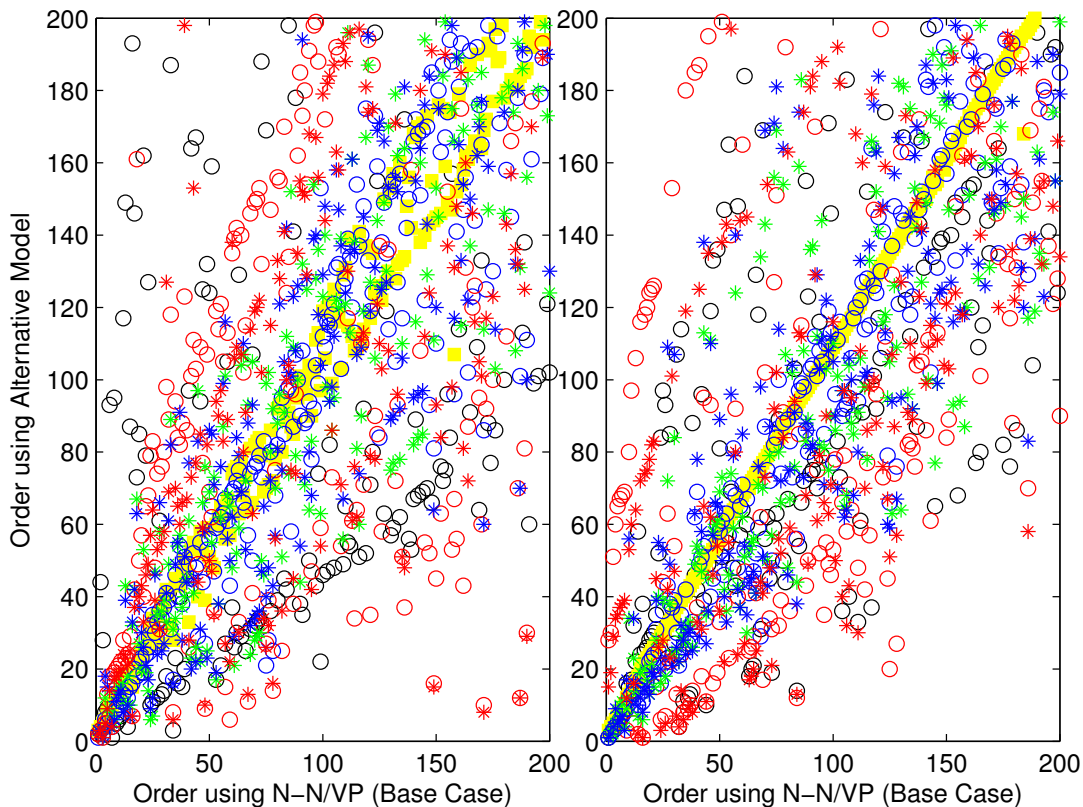


Fig. S6. Scatter plots showing the changes in the order of the top 200 compounds contributing to predicted SOA mass, compared to the base case (X-axis). SOA compositions used were from the standard (1.0/1.0/1.0) scenario; with:- panel A- $T = 273.15\text{K}$, $\%RH = 80$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core; panel B- $T = 303.15\text{K}$, $\%RH = 10$ and $3.0 \mu\text{grams.m}^{-3}$ involatile core. Key to the models is the same as in Figure 3 in the main paper:- Black circles:- Non Ideal; Yellow squares:- Hydrolysed; Blue circles:- SB-N/VP; Red circles:- JR-N/VP; Green asterisks:- N-MY; Blue asterisks:- SB-MY; Red asterisks:- JR-MY.

3 The use of SMARTS to identify groups for the Group Contribution Methods from molecular structures.

The software used to parse molecular structures into the groups required for the physical property estimation methods is written in Python. A Python interface has been developed for our selected chemical parsing software (OpenBabel)(O'Boyle et al., 2008). The Automated extraction of molecular substructure information is performed using the SMILES format (Simplified Molecular Input Line Entry System) for the target molecule. SMILES is a simplified chemical notation that allows a user to represent a two dimensional chemical structure in linear textual form. The notation is commonly employed in commercial and public software for prediction of chemical properties. It can be imported by most molecule editors for conversion into 2D/3D models and has a wide base of software support and extensive theoretical backing (see www.daylight.com). The OpenBabel chemical toolbox has the ability to filter and search molecular files using the SMARTS format (created by Daylight Chemical Information Systems, Inc alongside the SMILES format). SMARTS have been used, often in addition to descriptors derived from molecular modeling software, in the prediction of aqueous solubility of drug-like molecules (Hou et al., 2004; Klopman et al., 1992; Klopman and Zhu, 2001). Prediction of pure component vapour pressures requires development of bespoke SMARTS libraries. For this work, SMARTS libraries have been designed for a set of vapour pressure models using the 2742 compounds within the master chemical mechanism (MCM; Jenkin et al. (2003). Figure S7 shows the SMARTS required to identify some functional groups (Nannoolal primary groups see - Nannoolal et al. (2004)) that contribute to the structure of selected MCM compounds shown in Figure S8:-

Table S01:- SMARTS for some key Nannoolal primary groups

| Functional Group | Nannoolal Group | SMARTS |
|--------------------------|--------------------------------|--|
| A -COOH | NG_44 | [#6][CX3](=[OX1])[OX2;H1] |
| B -OOH | New group | [#6;!\$([CX3]=[OX1])][OX2][OX2;H1] |
| C -OH(primary) | NG_35 or NG_36 | SMARTS:- a:-[OX2;H1][CX4;H2,H3] b:-[OX2;H1;!\$(O[#6][#6,#7,#8][#6,#7,#8][#6,#7,#8][#6])][CX4;H2,H3;!\$(O[#6][#6,#7,#8][#6,#7,#8][#6,#7,#8][#6])] c:-[OX2;H1;!\$(O[#6][#6,#7,#8][#6,#7][#6][#6])][CX4;H2,H3;!\$(O[#6][#6,#7,#8][#6,#7][#6][#6])] d:-[OX2;H1;!\$(O[#6][#6,#7][#6][#6,#7,#8][#6])][CX4;H2,H3;!\$(O[#6][#6,#7][#6][#6,#7,#8][#6])] e:-[OX2;H1;!\$(O[#6][#6][#6][#6])([#6])([#6])][CX4;H2,H3;!\$(O[#6][#6][#6][#6])([#6])([#6])([#6])] |
| D -OH(vinyl) | assigned to - OH(sec) NG_34 | [OX2;H1;\$([OX2;H1][CX3]=[CX3])] |
| E -OH(tertiary) | NG_33 | [OX2;H1;\$([OX2;H1][CX4;H0])] |
| F -NO2(aliphatic) | NG_68 | [C][NX3](=[OX1])(=[OX1]) |
| G -ONO2 | NG_72 | [#6;!\$([CX3]=[OX1])][OX2][NX3](=[OX1])(=[OX1]) |
| H >C=O(aliphatic) | NG_51 | [C][CX3](=O)[C] |
| I -CHO(aliphatic) | NG_52 | [CX3;H1](=O)[C] |
| J -OO-(bridging) | NG_94 | [#6][OX2;R][OX2;R][#6] |

Fig. S7. SMARTS for some key Nannoolal primary groups

With regards to Figure S7, there are some specific items to consider for functional groups 'B', 'C' and 'D'.

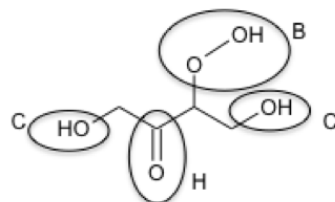
B:- Recursive SMARTS "!\$([CX3]=OX1)" is used to specify that the carbon holding the hydroperoxide group must not also be double bonded to an oxygen (to avoid hitting peroxyacids).

C:- It is easier enough to identify all primary alcohols (SMARTS a) but the Nannoolal method requires primary alcohols to be split between NG_35 (carbon chain of 5 or more atoms) and NG_36 (primary alcohols on a C4 or smaller chain) although the exact criteria for this split is not clear in the literature. In this work the allocation of primary alcohols is achieved using a set of five SMARTS. SMARTS b identifies whether the primary alcohol is on a carbon chain of 5 or more atoms. This chain has to be terminated by carbon atoms (which may bear functional groups that are not part of this count), but the intermediate atoms can be N or O as well as C. Hence (using SMILES notation) OCCCCO and OCCCCO would both have two alcohol groups belonging to NG_36 while OCCCCC, OCCOCC and OCCN(C)CC would have primary alcohols belonging to NG_35. The other three SMARTS account for the possible branching of this heavy atom chain:- thus OCC(C)(C) and OCN(C)C would both be NG_36 alcohols while OCC(C)(C)C and OCN(C)CC would be NG_35 alcohols.

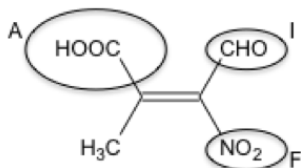
D:- Alcohol groups attached to a carbon-carbon double bond (vinyl alcohols) are not covered by the Nannoolal method. These SMARTS are used to identify vinyl alcohols which are then treated like secondary alcohols within the GCM.



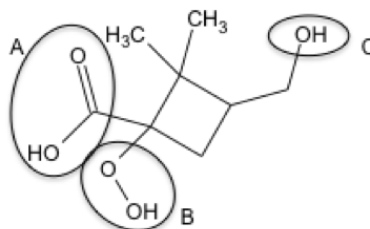
MCM 3442:- MALANHY
(After Hydrolysis)



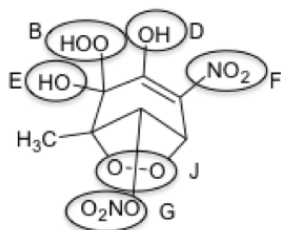
MCM2595-MVKOHBOOH



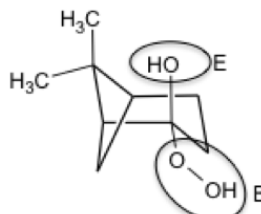
MCM 5276:- NC4MDCO2H



MCM 4834:- C812OOH



MCM 5240:- MNNCATCOOH



MCM 4610:- BPINAOOH

Fig. S8. Identification of some key functional groups for a small selection of MCM compounds. The functional groups are identified with letters and the associated SMARTS, for use with the Nannoolal group contribution methods, are given in Figure S4.

4 The correction to the Nannoolal Tb method for dicarboxylic acids.

In Barley and McFiggans (2010) it was shown that the slope of a vapour pressure curve with temperature, estimated using the Nannoolal method (Nannoolal et al., 2008), was generally accurate. The main source of error in the estimation of vapour pressure was in the calculation of the Tb value. In a further development of the Nannoolal vapour pressure equation Moller et al. (2008) recognized that strongly hydrogen bonding groups such as alcohols and carboxylic acids were not adequately represented in the Nannoolal method and included extra terms in the vapour pressure equation to handle multiple occurrences of these groups. It is not surprising that similar issues affect the estimation of Tb values and this is shown by the discrepancies between predicted Tb values for dicarboxylic acids by the Nannoolal Tb method (Nannoolal et al., 2004) and experimentally measured vapour pressures for a range of linear dicarboxylic acids.

Fifty four sets of experimental vapour pressure data for dicarboxylic acids from 12 sources were obtained from the literature. In most cases the experimental data were for solid state samples so experimental latent enthalpies of fusion were required to correct the solid state vapour pressure values to sub-cooled liquid values (Barley and McFiggans, 2010; Prausnitz et al., 1986). The vapour pressure values were then extrapolated using the Nannoolal vapour pressure equation up to one atmosphere to obtain an experimentally derived estimate of the normal boiling point for the diacid. These values are plotted, along with the estimated Tb values by the Nannoolal method, in Figure S9:-

Although there is considerable scatter in the values it is clear that the Nannoolal method is substantially underpredicting the Tb for dicarboxylic acids, especially those of low carbon number. From these data ΔT (the difference between the experimentally derived Tb value and the value estimated by Nannoolal) could be calculated and then correlated as a function of carbon number. The correction to the Nannoolal Tb

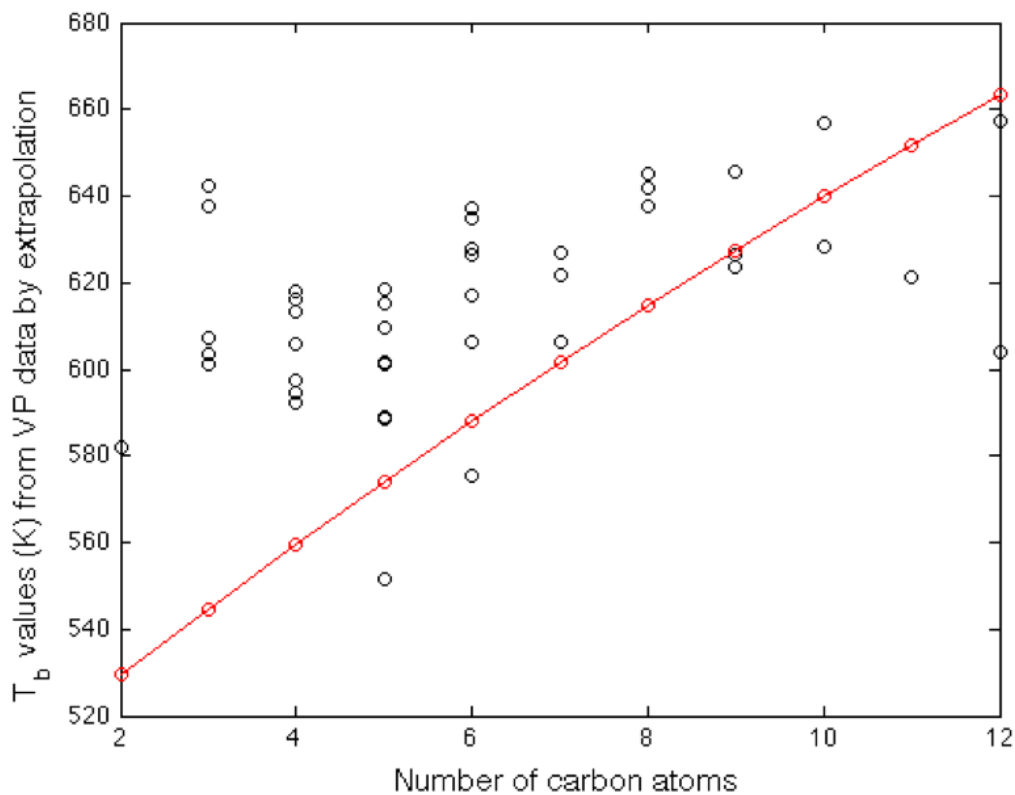


Fig. S9. A comparison of the experimentally derived normal boiling point values for a range of linear dicarboxylic acids with the boiling points estimated by the Nannoolal T_b method (red line).

estimation methods for dicarboxylic acids is:-

$$\Delta T = -9.2169C + 84.11 \quad (1)$$

where C is the number of carbon atoms in the diacid. This correction is about +50K for a C4 diacid, +20K for a C7 and becomes negative for a C10 molecule. It should not be used outside the range C3-C12.

References

- Barley, M. H. and McFiggans, G.: The critical assessment of vapour pressure estimation methods for use in modelling the formation of atmospheric organic aerosol, *Atmospheric Chemistry and Physics*, 10, 749–767, 2010.
- Hou, T. J., Xia, K., Zhang, W., and Xu, X. J.: ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach, *Journal of Chemical Information and Computational Science*, 44, 266–275, 2004.
- Jenkin, M. E., Saunders, S. M., Wagner, V., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds, *Atmospheric Chemistry and Physics*, 3, 181–193, 2003.
- Klopman, G. and Zhu, H.: Estimation of the aqueous solubility of organic molecules by the group contribution approach, *Journal of Chemical Information and Computational Science*, 41, 439–445, 2001.
- Klopman, G., Wang, S., and Balthasar, D. M.: Estimation of Aqueous Solubility of Organic Molecules by the Group Contribution Approach. Application to the Study of Biodegradation, *Journal of Chemical Information and Computational Science*, 32, 474–482, 1992.
- Moller, B., Rarey, J., and Ramjugernath, D.: Estimation of the vapour pressure of non-electrolyte organic compounds via group contributions and group interactions, *Journal of Molecular Liquids*, 143, 52–63, 2008.
- Nannoolal, Y., Rarey, J., Ramjugernath, D., and Cordes, W.: Estimation of pure component properties Part 1. Estimation of the normal boiling point of non-electrolyte organic compounds via group contributions and group interactions, *Fluid Phase Equilibria*, 226, 45–63, 2004.

O'Boyle, N. M., Morley, C., and Hutchison, G. R.: Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit., *Chemistry Central journal*, 2, 5, doi:10.1186/1752-153X-2-5, 2008.

Prusnitz, J. M., Lichtenhaler, R. N., and de Azevedo, E. G.: *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, Englewood Cliffs, N J, USA, 2nd edn., 1986.