An Assessment of Vapour Pressure Estimation Methods†

Simon O'Meara,* Alastair Murray Booth,* Mark Howard Barley,* David Topping* and Gordon McFiggans*

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Laboratory measurements of vapour pressures for atmospherically relevant compounds were collated and used to assess the accuracy of vapour pressure estimates generated by seven estimation methods and impacts on predicted secondary organic aerosol. Of the vapour pressure estimation methods that were applicable to all the test set compounds, the Lee-Kesler [Reid *et al., The Properties of Gases and Liquids*, 1987] method showed the lowest mean absolute error and the Nannoolal et al. [Nannoonal *et al., Fluid Phase Equilibria*, 2008, **269**, 117-133] method showed the lowest mean bias error (when both used normal boiling points estimated using the Nannoolal et al. [Nannoolal *et al., Fluid Phase Equilibria*, 2004, **226**, 45-63] method). The effect of varying vapour pressure estimation methods on secondary organic aerosol (SOA) mass loading and composition was investigated using an absorptive partitioning equilibrium model. The Myrdal and Yalkowsky [Myrdal and Yalkowsky *Industrial & Engineering Chemical Reserach*, 1997, **36**, 2494-2499] vapour pressure estimation method using the Nannoolal et al. [Nannoolal *et al., Fluid Phase Equilibria*, 2004, **226**, 45-63] normal boiling point gave the most accurate estimation of SOA loading despite not being the most accurate for vapour pressures alone.

1 Introduction

Aerosol influence the Earth's radiative budget through the scattering and absorption of radiation[1]. Additionally they can adversely affect health[2,3](Pope and Dockery, 2006; Heal et al., 2012), and atmospheric trace gas constitution by providing reagents for chemical reactions and providing sites for chemical reaction[4].

Quantification of these effects is limited, in part, by our inability to predict aerosol composition[1,5] and microphysics. This is complicated by the large number of organic compounds present in the atmosphere, which explicit automatic generation mechanism techniques estimate to be or the order 10^6[6]. The mass fraction of aerosol formed of organic material varies by location, but it is frequently significant and can be dominant[7,8]. Once emitted, volatile organic compounds (VOCs) can undergo chemical reactions, leading to products of generally lower volatility and greater propensity to condense.

Organic products with volatilities that allow fractions to exist in both the condensed (particle) and gas phase are referred to as semi-volatile organic compounds (SVOCs). In this paper SVOCs are defined as compounds able to partition (under any condition) between 99.9 % and 0.1 % into the condensed phase, the remainder being in the gas phase (based on volatility considerations only).

In addition to the oxidation of VOCs, SVOCs may also enter the gas phase by the evaporation or sublimation of compounds already in the particulate phase. Examples include chemicals within particles being warmed and/or diluted[9]. The volatility of a given SVOC, given by its saturation vapour pressure (p^o), is one thermodynamic property determining the condensable fraction. SVOCs can constitute a major fraction of total organic aerosol[8], therefore (p^o) values are required to accurately estimate particulate composition and mass loading when using a mechanistic approach.

Fig. 1 illustrates how the condensable fraction of a compound varies with p^o, the particulate mass and temperature (when p^o is the only thermodynamic property considered). The C^* metric in Fig. 1 is the effective saturation coefficient defined in[9] and calculated by:

$$C_i^* = \frac{10^6 M_i p_i^o}{RT} \tag{1}$$

where M_i is the molar mass of compound i, R is the ideal gas constant (8.2057x10-5 m^{-3} atm mol^{-1} K^{-1}) and T is the temperature. $\log_{10}(C^*)$ is used as the abscissa in the volatility basis set (VBS) method of partitioning estimation, in which compounds are typically aggregated into unit bins[9]. This method can be constrained using experimental data from chamber experiments or as a technique for grouping

---

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

Centre for Atmospheric Sciences, School of Earth, Atmospheric and Environmental Sciences, University of Manchester, Manchester, M13 9PL, UK. Tel: 01613063954; E-mail: g.mcfiggans@manchester.ac.uk

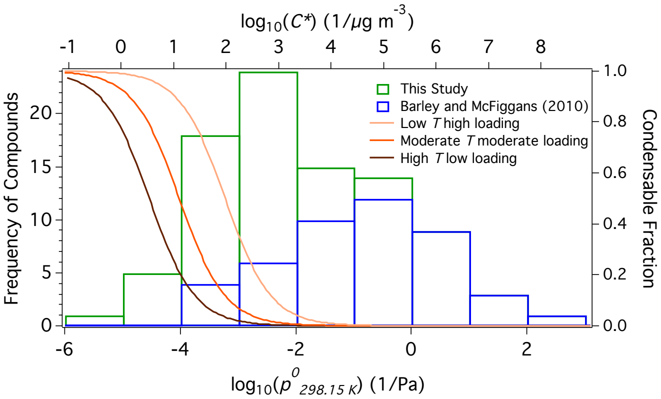

Fig. 1 Histogram (left axis) for the vapour pressure (bottom axis) of compounds in this study and those used in [10]. Orange curves correspond to the right axis and represent the condensable fraction change with effective saturation coefficient (top axis), where 1 means entirely partitioned into the condensed phase. Condensable fractions were obtained for different scenarios, varying by temperature (T) and particulate loading: Low temperature = 273.15 K, low loading = 3.0 μg m^{-3}, moderate temperature = 293.15 K, moderate loading = 12.0 μg m^{-3} and high temperature = 313.15 K, high loading = 96 μg m^{-3}.

more detailed single component representations of condensation/evaporation. It is included in Fig. 1 since it is a familiar measure, however, as discussed in Barley *et al.* [11] a molar based approach is more reliable. Molar based calculations are described in eqs. 13 and 14 below. For Fig. 1 an average molar mass of 200 g mol^{-1} was assumed.

For estimating particulate mass loading in atmospheric scenarios, transport models may use the VBS. This semi-empirical approach, however, relies on a simplified treatment of the chemical composition of SVOCSs. The alternative mechanistic approach is to predict the chemicals present through a chemical degradation model followed by estimation of their vapour pressures using a p^o estimation method such as one of those assessed below[12]. Assessment of the sensitivity of such a mechanistic approach to estimated p^o and other inputs is required to quantify uncertainties in particulate loading and prioritise development of model components. For instance, loadings were noted to be particularly sensitive to vapour pressures estimates in McFiggans *et al.* [13].

Laboratory measurements of saturation vapour pressure have been made by several methods[14,15]. Due to their large number, measurement of the vapour pressure of all atmospheric organic compounds is currently impractical, necessitating p^o estimation methods. Multiple estimation methods have been proposed, varying structurally and by their parameterisations. Seven estimation methods, chosen for their relatively high accuracy (as shown in prior assessments), are assessed here for accuracy.

The assessment of vapour pressure estimation methods by Barley and McFiggans [10] compared estimates against vapour pressure measurements. The current study follows an identical approach but doubles the number of test set compounds from Barley and McFiggans [10] (90 vs. 45). The average measured p^o of test set compounds is lowered, as shown in Fig. 1 (average p^o at 298.15 K here is 2.7 Pa compared to 5.3 Pa in the earlier study).

Fig. 1 shows that even at a comparatively high particulate loading and low temperature (conditions that act to increase the condensable fraction of a given compound), the majority of compounds used in the test set of Barley and McFiggans [10] would not partition $>$10 % into the condensed phase. In contrast, approximately half of the compounds in the new study would partition at least 20 % into the condensed phase. The test set of compounds used here is therefore more representative of SVOCs than the previous assessment.

Camredon and Aumont [16] and Mbienda *et al.* [17] assessed vapour pressure estimation methods using larger test sets of measurements than our study, however, some of their measurements are collated from secondary sources, such as data books. Such sources can be unclear as to the primary source of the measurement values and whether an extrapolation has been applied from a different temperature, thereby introducing uncertainty. Consequently, all measurements used here are from the primary literature. Furthermore, as with the Barley and McFiggans [10] study, the majority of measurements were above 10^{-1} Pa. For Camredon and Aumont [16] only 18 of a total 850 measurements fall below this pressure, and there is a similar proportion in the Mbienda *et al.* [17] study, therefore for testing compounds relevant to SVOC with low p^o the current study is advantageous. Unlike these past assessments, the current study also evaluates the sensitivity of SOA loading and composition to vapour pressure estimation methods.

Both Camredon and Aumont [16] and Mbienda *et al.* [17] assessed the Lee-Kesler [18] and Myrdal and Yalkowsky [19] methods, which are also assessed here. Camredon and Aumont [16] showed these two methods to have similar estimation accuracy, but Myrdal and Yalkowsky [19] showed greater potential for accurate estimates at lower p^o. Mbienda *et al.* [17] reported that the accuracy for both methods tends to decrease with increasing number of oxygenated functional groups per compound, but that this trend was stronger for Lee-Kesler.

The estimates of several estimation methods were compared against those of the EVAPORATION method in Compernolle *et al.* [20]. The method of Nannoolal *et al.* [21] had the strongest agreement, and since this shows comparatively high accuracy in the Barley and McFiggans [10] assessment, this suggests EVAPORATION could also produce accurate estimates. Here EVAPORATION[20] is assessed against measurements for the first time (to the author's knowledge).

This study reports the results of the comparison of estimated

p^o (from seven estimation methods) against measured p^o for a test set of compounds. To assess the wider significance of these results, measured and estimated p^o are also applied to an absorptive partitioning model. We present the SOA loading and composition sensitivity to varying p^o estimates, and, through comparison with measured vapour pressure values, identify the estimation method predicting the most accurate SOA loading and composition.

2 Method

2.1 Test Set Compiliation.

Laboratory measurements of vapour pressure for 90 compounds were collated. Recent studies have presented evidence of laboratory and ambient particulates taking a non-crystalline solid state[22,23], though the actual phase state of complex mixtures remains debatable. We use the sub-cooled liquid vapour pressure for two reasons. Estimation methods for a non-crystalline solid state are not, as far as the authors are aware, specifically presented in the literature. Consequently, the current modeling approach to highly viscous mixtures is to use the sub-cooled liquid p^o reference state and account for the influence of viscosity through diffusion limitations in the condensed phase. Therefore, the presented assessment of sub-cooled liquid p^o methods is relevant to SOA formation models. There remains substantial debate around the phase of organic components in particulates. According to thermodynamic theory[24] mixing of different compounds in the condensed phase depresses the melting point of each species, with the degree of depression correlated with the number of species. Given that the estimated number of atmospheric organic compounds contributing to the condensed phase might be of the order 10^5 [6] there is theoretical support for the organic component of particles to be in a liquid-like state. Furthermore, a liquid-like state has been observed under laboratory conditions[25]. More work is needed to decipher the true phase state and subsequent influence on gas/particle mass transfer.

Where necessary solid state vapour pressures measurements were converted to sub-cooled liquid vapour pressures using the equation of Prausnitz *et al.*[26]:

$$\ln(p_l^o) = \ln(p_s^o) + \frac{\Delta H_{fus}}{RT_m}\left(\frac{T_m}{T}-1\right) - \frac{\Delta c_{p,sl}}{R}\left(\frac{T_m}{T}-1\right) + \frac{\Delta c_{p,sl}}{R}\ln\left(\frac{T_m}{T}\right) \quad (2)$$

where p_l^o and p_s^o are the sub-cooled liquid and solid vapour pressures respectively, the latter obtained at temperature T. ΔH_{fus} is enthalpy of fusion, R is the ideal gas constant (8.314 J $\mathrm{mol^{-1}K^{-1}}$), T_m is the melting temperature and $\Delta c_{p,sl}$ is the heat capacity change between the solid and liquid state. If $\Delta c_{p,sl}$ measurements were not available the assumption $\Delta H_{fus}/T = \Delta S_{fus} = \Delta c_{p,sl}$ was used. Sources of ΔH_{fus}, $\Delta c_{p,sl}$ and T_m are given in Table 1 of the supplementary material†. Measurements were selected if they met the following criteria:

1. Compounds were required to have at least one non-hydrocarbon functional group, since oxidation is the major mechanism by which VOCs are converted to SVOCs[27,28].
2. The non-hydrocarbon functional group(s) borne by compounds had to be reported as having been observed in the atmosphere. The selected atmospherically relevant functional groups and their associated publications are listed in Table 1.
3. Measurements had to be below 10^3 Pa, since compounds with vapour pressures above this are expected to be too volatile to contribute to the condensed phase.
4. For measurements of solid vapour pressure, there had to be a measurement of ΔH_{fus} and T_m available in the primary literature to apply the conversion to sub-cooled p^o (eq. 2).

Table 1 The functional groups included in the test set.

Functional Group	Number of compounds with group	Publication supporting ambient presence
Carboxyl	51	Pietrogrande *et al.*[29]
Hydroxyl	29	Decesari *et al.*[30]
Amine	11	Beddows *et al.*[31]
Ketone	7	Hamilton *et al.*[32]
Aldehyde	5	Hamilton *et al.*[32]
Nitro	9	Rogge *et al.*[33]
Ester	3	Helmig *et al.*[34]
Ether	7	Kalberer *et al.*[35]
Methoxy	12	Decesari *et al.*[30]
Nitrate	1	Day *et al.*[36]

The selected measurements and their associated publications are presented in Table 1 of the supplementary material. To describe the test set compounds, their average carbon molecule oxidation state[37] and number of carbon atoms per compound are presented in Fig. 2. Also in this figure, are the sub-cooled liquid vapour pressures at 298.15 K (extrapolated where necessary). Lower volatility compounds are expected toward the top left corner of the Fig. 2 plot, i.e., high oxidation state and high carbon number. Although existent compounds do fit this space[37], their vapour pressures have not been measured, owing to the difficulty of their measurement[14,15,38]. The reader is referred to the original publications of measurements, given in Table 1 of the supplementary material, for a description of measurement methods.

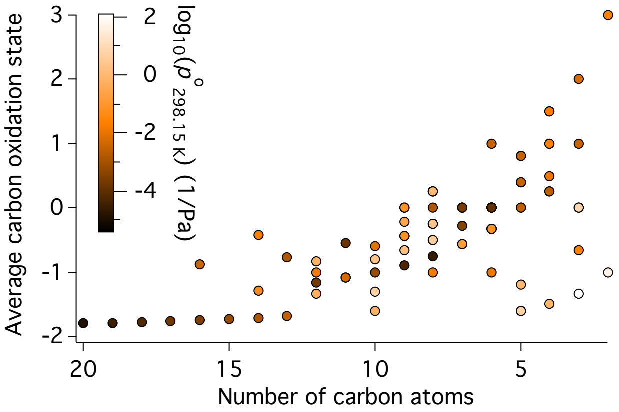

Fig. 2 The average carbon oxidation state and carbon number for each compound in the test set. Colours represent the sub-cooled liquid vapour pressure at 298.15 K, as described in the colour scale.

Unfortunately, varying fractions of test set measurements have been used in the parameterisation of the different estimation methods, introducing some bias. As shown in Table 1 and Table 1 of the supplementary material, dicarboxylic acids constitute a substantial fraction of the test set. Because of this, the EVAPORATION method, in particular, is likely to have a favourable bias, since this was created most recently, and was therefore able to use the recent widespread availability of p^o measurements for dicarboxylic acids, most of which are also part of the assessment test test. 13 of the 14 primary literature sources of dicarboxylic acid p^o measurements used for EVAPOARTION have been published since 2000. The methods of Lee-Kesler and Myrdal and Yalkowsky, however, were produced before then, and are therefore expected to perform less well for this class of compound. A further limitation of the test set concerns its representativeness of the atmosphere. Ideally all functional groups observed in ambient SVOCs would be represented. Table 1 shows that although a reasonably representative range of atmospherically relevant functional groups are covered, peroxyacyl nitrate, hydroperoxide and percarboxylic acid groups are not, resulting from a scarcity of p^o measurements. Preferably we would be able to assess against more measurements of nitrates, since these can constitute a substantial fraction of organics in the condensed phase[36]. The high reactivity of nitrate compounds, however, often makes their handling unsafe and therefore greatly limits vapour pressure measurements. This is also an important factor restricting measurement of peroxyacyl nitrate, hydroperoxide and percarboxylic acid compounds, in addition to their thermal instability. Further welcome additions to the test set to improve atmospheric representativeness would be: methanesulfonic acid[39], organic salts[40,41], organosulfates[42] and compounds where the only non-hydrocarbon functional group is amines[41] or ketones/aldehydes[35]. Generally, multifunctional compounds with low vapour pressures ($<1\text{x}10^{-3}$) other than substituted dicarboxylic acids (which have had comparatively large numbers of measurements) would be useful additions.

With regard to the ratios of numbers of compounds bearing particular functional groups, observational studies show there to be large variation in space and time[43,44]. Furthermore, much of the condensed organic component is yet to be identified at a molecular level[8], limiting judgment on the atmospheric relevance of the test set with regard to the proportions of compounds with certain functional groups. The assessment was not only conducted on the entire test set, but also for low volatility compounds, with $p^o<10^{-2}$ Pa, only. As can be seen from 1, below this vapour pressure, notable fractions of compounds will condense, making them most relevant to secondary organic aerosol. Compounds with particularly poor or good vapour pressure estimation were also evaluated in greater detail.

p^o measurements were taken at a range of temperatures between 263 and 405 K. At values outside the ambient temperature range, relevance to the atmosphere is somewhat decreased. Furthermore, the difference between estimated and experimental p^o at the measurement temperature and at ambient temperatures is likely to vary. However, through extrapolating p^o values to ambient temperature using, for example, an Antoine equation, error is introduced by uncertainty in the equation parameters. For the assessment of compound vapour pressure estimation there was no benefit in extrapolation to ambient temperature, as it would have increased the uncertainty in the measurement values, thereby complicating the source of errors (i.e., from measurement values or estimation methods). The majority of compounds (66) did have measured vapour pressures at ambient temperatures.

For the same reason, for the low volatility subset of compounds, although they were selected based on their vapour pressure at 298.15 K (extrapolated to where necessary), their vapour pressures at measurement temperatures were used for the assessment. In contrast, where necessary, vapour pressures were extrapolated to 298.15 K for the SOA loading assessment, in order to best simulate ambient conditions.

2.2 Estimation Methods.

The seven estimation methods assessed were chosen for having shown comparatively high accuracy in previous assessments[10,45].

The majority of methods tested take a group contribution approach, where functional groups (e.g. -CH3) are first defined, followed by estimation of their contribution to p^o through a minimisation procedure using a set of measured p^o

for a variety of compounds (often referred to as the training set). The constitution of the training set and subsequent parameterisation is therefore of fundamental importance to the accuracy of the resulting estimates (this point is also true for the non-group contribution methods). To estimate p^o, group contributions are multiplied by the frequency of the corresponding functional group in a given compound and these products are summed. The functional groups of a given method determine the compound classes it is applicable to.

Four of the estimation methods: Nannoolal *et al.*[21], Myrdal and Yalkowsky[19], Moller *et al.*[46] and Lee-Kesler[18] are combined methods, requiring T_b as input. T_b was found using the group contribution estimation methods of Nannoolal *et al.*[47] and Stein and Brown[48]. The former had a correction applied to improve accuracy for small dicarboxylic acids, as described in the supplementary material of Barley *et al.*[49]. The popular T_b method of Joback and Reid[50] was not used as it has been shown to produce comparatively low accuracy estimates[10,48].

The input T_b, however, is a potential source of error. For methods that extrapolate p^o from T_b, the magnitude of propagated error increases with increasing difference between T_b and the temperature to be estimated at. If any error in the p^o estimation, however, acts to compensate the propagated error in T_b a reasonably accurate result can be obtained. This has been observed when the Myrdal and Yalkowsky[19] p^o estimation method is combined with the Joback and Reid[50] T_b estimation method, with the latter's tendency to overestimate T_b often negated by the former's tendency to underestimate the slope of the vapour pressure line[10,51]. Nevertheless, this combination was not tested here, since the assessment was not concerned solely with identifying the estimation method generating the highest accuracy p^o, but the one that best modeled the dependence of p^o on temperature.

The Nannoolal *et al.*[21] method was found to give comparatively high accuracy p^o in the Barley and McFiggans[10] assessment. It relies on the normal boiling temperature (T_b) of a compound from which it extrapolates the p^o vs. T curve. The curve's gradient is estimated using a group contribution approach. For compound i:

$$\log_{10}(p^o_{i,l}) = (4.1012 + dB)\left(\frac{T_r - 1}{T_r - \frac{1}{8}}\right) \tag{3}$$

where dB is the group contribution and $T_r = T/T_b$. dB is found using group contribution values. In addition to functional groups, some steric effects and group interactions are considered in estimating dB; these are often referred to as second-order effects.

Moller *et al.*[46] noted the Nannoolal *et al.*[21] method performed comparatively poorly for aliphatic compounds containing carboxyl and alcohol groups. They applied a correction term to resolve this. Additionally, the training set of experimental results was expanded from Nannoolal et al., generating a different parameterisation:

$$\ln(p^o_{i,l}) = (9.42208 + dB)\left(\frac{T - T_b}{T - C}\right) + D'\ln\left(\frac{T}{T_b}\right) \tag{4}$$

where the second term is the correction term and only used for compounds containing carboxyl and alcohol groups. C is given by:

$$C = 2.65 + \frac{T_b^{1.485}}{135} \tag{5}$$

Despite these modifications Compernolle *et al.*[45] and Booth *et al.*[51] observed that this method estimated unrealistic vapour pressures for molecules with high numbers of hydrogen bonding groups.

Of the group contribution methods assessed here, Nannoolal *et al.* and Moller *et al.* have the highest number of groups: 132 for the former and 212 for the latter. Although this complexity has the potential to enable better accuracy through greater specificity, groups trained on too few measurements are liable to over-fitting. To an extent, this problem has been mitigated by the relatively large size of the training set for these methods: approximately 1660 and 2330 compounds for the Nannoolal *et al.* and Moller *et al.* methods, respectively.

Like the previous two, the Myrdal and Yalkowsky[19] method is a combined method, requiring T_b for input. It is not a group contribution method however, rather intermolecular bonds and molecular flexibility are considered:

$$\log_{10}(p^o_{i,l}) = -\frac{(86.0 + 0.4\tau + 1421(\mathrm{HBN}))(T_b - T)}{19.1T} + \frac{(-90.0 - 2.1\tau)}{19.1}\left(\frac{T_b - T}{T} - \ln\left(\frac{T_b}{T}\right)\right) \tag{6}$$

where HBN denotes the hydrogen bond number and τ is the effective torsional bond number. Both Compernolle *et al.*[20] and Barley and McFiggans[10] observed a systematic increase in overestimation with decreasing volatility for this method. However, a comparatively low standard deviation around its line of best fit was also noted.

The method of Capouet and Muller[52] uses the vapour pressure of the parent hydrocarbon $(p^o_{hc,l})$ (the molecule left when all oxygenated functional groups are replaced by a H atom):

$$\log_{10}(p^o_{i,l}) = \log_{10}(p^o_{hc,l}) + \sum_{k=1}^{n} N_{k,i}(B + CT) \tag{7}$$

where N is the number of group k in the compound and B and C are group contributions. n is the total number of groups. Where possible parent hydrocarbon vapour pressures were found from the primary literature (the sources are listed beside their compounds in Table 1 of the supplementary material). Where literature values could not be found $p^o_{hc,l}$ was estimated using the Lee-Kesler method[18], following the recommendation of Capouet and Muller[52]. This method was only applicable to 9 compounds in the Barley and McFiggans[10] assessment due to the limited number of functional groups it accounts for. For these 9 it showed a comparatively low standard deviation and bias. In this study the method is applicable to 42 compounds, allowing a more robust assessment of its accuracy.

Compernolle *et al.*[20] proposed the EVAPORATION method:

$$\log_{10}(p^o_{i,l}) = A_i + \frac{B_i}{T^{1.5}} \tag{8}$$

where A and B are found from group contributions. 788 compounds were used in the training set of EVAPORATION. Like Nannoolal *et al.*[21] and Moller *et al.*[46] it attempts to account for second-order effects, though it uses 5 second-order groups compared to approximately 60 in Moller *et al.*[46]. Compounds with relevance to the atmosphere were targeted in the minimisation process used to quantify group contributions. This is not the case for all the methods described here, for example Nannoolal *et al.*[21] was produced with a view to applications in chemical engineering, though its functional groups cover those observed in the atmosphere very well. Several functional groups present in our test set compounds are not covered by the EVAPORATION method, including aromatics, thereby restricting it to 46 compounds. EVAPORATION has not been assessed against measured p^o other than in its original publication where it was shown to have a strong correlation to the Nannoolal *et al.*[21] method. This indicates potential to produce good estimates since Nannoolal *et al.*[21] demonstrated comparatively high accuracy in the Barley and McFiggans[10] assessment.

The SIMPOL.1 method by Pankow and Asher[53] is similar to EVAPORATION, though with a more complex temperature dependence:

$$\log_{10}(p^o_{i,l}) = A(T) + \sum_{k=1}^{n} N_{k,i}\left(\frac{C_{1,k}}{T} + C_{2,k} + C_{3,k} + C_{4,k}\ln(T)\right) \tag{9}$$

where A is constant (at a given temperature) across all compounds. C values are group contributions for group k. Halogen containing groups, nitrile and carbonate groups are not included in SIMPOL.1, making it applicable to 81 of our test set compounds.

Although SIMPOL.1 has greater degrees of freedom compared to EVAPORATION, its training set contained less compounds (272), making it more liable to over-fitting. Despite showing relatively inaccurate p^o in Barley and McFiggans[10] SIMPOL.1 continues to be a common choice of estimation method in studies considering SVOCs (e.g. Fry and Sackinger[54]), therefore it was chosen for further testing here.

The corresponding states method of Lee-Kesler[18] takes critical temperature (T_c) and pressure (p_c) as input:

$$\ln(p^o_{i,r}) = f^{(0)}(T_r) + \omega f^{(1)}(T_r) \tag{10}$$

where

$$f^{(0)} = 5.92714 - \frac{6.09648}{T_r} - 1.28862\ln(T_r) + 0.169347(T_r)^6 \tag{11}$$

and

$$f^{(1)} = 15.2518 - \frac{15.6875}{T_r} - 13.472\ln(T_r) + 0.43577(T_r)^6 \tag{12}$$

T_r is the reduced temperature: T/T_c, p_r is the reduced pressure: p/p_c. ω is the acentric factor, which depends on the critical properties as well as the boiling point. Critical properties were estimated using the group-contribution method of Ambrose[18] using T_b estimated by Nannoolal *et al.*[47] as primary input. Reasonable performance in comparison with other methods was observed for Lee-Kesler in Barley and McFiggans[10], though with a poorer fit to measured values than Nannoolal *et al.*[21]. Booth *et al.*[51], however, observed that Nannoolal *et al.*[21] poorly estimated p^o for compounds with 3 oxygenated groups or more. The average number of oxygenated functional groups per compound is increased in this study over Barley and McFiggans[10], giving a more representative sample of semi-volatile organics and a more robust test of estimation methods.

Table 2 lists all estimation methods and the number of test set compounds each could be applied to given the functional groups they covered. The full 90 compounds will be referred to as the complete test set, while the 42 compounds, which all methods could be applied to, will be referred to as the reduced test set. Similarly, for the subset of compounds with $p^o_{298.15K}$ <10^{-2} Pa, Nannoolal *et al.*[21], Lee-Kesler, Myrdal and Yalkowsky[19] and Moller *et al.*[46] were applicable to all 50 compounds fitting this criteria, while SIMPOL.1, EVAPORATION and Capouet and Muller[52] were only applicable to 49, 34 and 33 compounds of this subset, respectively.

2.3 SOA Model.

The molar based absorptive partitioning model of Pankow[55] was used to estimate SOA mass loading using measured and estimated vapour pressures. Central to the model is calculation

Table 2 The assessed estimation methods and the number of test set compounds they were applicable to

Estimation Method	Applicable Compounds
Nannoolal *et al.*[21]	90
Moller *et al.*[46]	90
Myrdal and Yalkowsky[19]	90
Capouet and Muller[52]	42
Compernolle *et al.*[20]	46
Pankow and Asher[53]	81
Lee-Kesler[18]	90

of the condensable fraction of a given compound (i), called the partitioning coefficient:

$$\xi_i = \left(1 + \frac{C^*_{Mi}}{C_{MOA}}\right)^{-1} \tag{13}$$

where C^*_{Mi} is the molar based effective saturation coefficient, found by removing M_i from eq. 1. C_{MOA} is the total molar concentration (μmol m^{-3}) of condensed material and non-volatile mass:

$$C_{MOA} = \frac{\sum_i C_i \xi_i + C_{nv}}{\overline{M}} \tag{14}$$

where C_i and C_{nv} are the mass concentrations of SVOCs (including water) and the non-volatile part respectively. $\overline{M}$ is the average molar mass of material in the condensed phase. Eqs. 13 and 14 were run iteratively using a first-guess for the C_{MOA} value to initialise the calculation.

Similar studies of modeled SOA loading employ chemical mechanism models to estimate the atmospheric compounds present. Such studies rely on estimation methods for compound vapour pressures since measured vapour pressures for all the compounds predicted to be present are not available. There is no reason the values used in those studies are correct, as we have shown below in the comparison with pure component values. In addition, as other studies have also shown, sensitivity is dependent on the gas phase mechanism used and the number of compounds included in those schemes[49,56]. Here, compounds of the complete and reduced test sets that had measured vapour pressure values were used as two sets of SVOCs input to the partitioning model. Although this provides a simplified model of atmospheric particulate production, our ability to estimate loading using measured and estimated vapour pressures allows the error in loading when using estimation methods to be quantified. This is a major advantage over studies without measured vapour pressures. For this part of the assessment measured p^o were extrapolated to 298.15 K (where necessary). Concentrations of each compound were fitted to observations, providing the following constraints:

1. when measured p^o values were used the total aerosol mass concentration was either 3.0 or 96.0 μg m^{-3}, simulating a remote and a polluted environment, respectively[57]
2. compounds were segregated into unit bins of $\log_{10}(C^*)$ depending on their measured p^o. The relative amounts of aerosol mass per bin were fitted to the observations of total organic aerosol mass concentration in the field study of Cappa and Jimenez[58] (their Fig. 5a).

The resulting concentrations for compounds in each volatility bin for both the remote and polluted environment are shown in Table 3.

The total number of organic compounds in the atmosphere is several orders of magnitude higher than the number in the test set. However, the test set compounds have been fitted with concentrations to reproduce observed concentrations of organic aerosol. Consequently, the fitted concentrations are unrealistically high. Nevertheless, this approach provides the most realistic interpretation of SOA formation using available measurements of p^o for compounds pertinent to SVOCs.

Cappa and Jimenez[58] found that the global minimum C^* for SVOCs is 10^{-3} μg m^{-3} by considering the lowest aerosol mass loading and highest temperatures realistic for Earth (conditions acting to promote evaporation of SVOC). This was taken to be the minimum volatility for SVOCs in the current study, however, the minimum C^* of the test set compounds was 10^{-1} μg m^{-3}. To fill the 10^{-3} and 10^{-2} μg m^{-3} C^* bins a concentration was assigned to each, fitted as for the test set compounds. A molar mass of 200 g mol^{-1} was assumed for these bins.

The concentrations of test set compounds were held constant when p^o estimation methods were applied. If test set compounds had a p^o predicted by an estimation method that placed them in the lowest two volatility bins their concentration was summed with the concentrations assigned to these bins in the measured p^o case.

To represent the non-volatile fraction a non-volatile mass concentration was introduced with an assumed molar mass of 200 g mol^{-1}. The ratio of non-volatile to condensed organic aerosol mass concentrations was set to 1:1 for the measured p^o, and the resulting non-volatile mass was maintained throughout the model runs using estimated p^o. This ratio is typical of observations, though it can vary substantially[57,58]. Relative humidities between 20-90 % were tested to determine the effect on SOA loading for a given estimation method. The variation was found to be negligible in comparison to the variation between estimation methods. A single relative humidity of 60 % was therefore used to obtain the results shown below.

The partitioning model had several assumptions: ideality, no diffusion limitations and no consideration of phase separation, therefore volatility was the sole determinant of gas/condensed partitioning. Despite these limitations the

Table 3 The total (sum of the condensed and gas phase) concentration allocated to each compound in a $\log_{10}(C^*)$ bin. Bins are given by their central value, e.g., -2.5 represents the range -3 $\leq \log_{10}(C^*) <$-2.

Number of compounds in test set	Total concentration per $\log_{10}(C^*)$ bin (μgm^{-3}) (x10^{-1})									
	2	3	4	5	6	7	8	9	10	11
90	3.7	4.0	4.3	0.79	0.46	0.66	3.9	8.7	35	290.0
42	3.6	3.8	4.1	1.1	6.4	1.1	41.0	27.0	99.0	820.0
	Number of compounds per $\log_{10}(C^*)$ bin (μgm^{-3})									
90	1	1	1	7	19	27	11	13	9	3
42	1	1	1	5	13	16	1	4	3	1

model is thought to be sufficient to gain a primary indication of the sensitivity of aerosol loading and composition to vapour pressure.

3 Results and Discussion

3.1 Vapour Pressure Estimates Compared to Measurements.

The mean absolute error (MAE) and mean bias error (MBE) for each method are presented in Table 4. In order to weight errors equally between compounds with varying numbers of measurements, the average error was first found for a given compound (averaging over multiple measurements (j)), before the average over all compounds (i) was found:

$$\mathrm{MAE} = \frac{1}{n_i}\sum_{i=1}^{n_i}\left(\frac{1}{n_j}\sum_{j=1}^{n_j}\left|\log_{10}\frac{p^o_{est}}{p^o_{exp}}\right|\right) \tag{15}$$

and

$$\mathrm{MAE} = \frac{1}{n_i}\sum_{i=1}^{n_i}\left(\frac{1}{n_j}\sum_{j=1}^{n_j}\log_{10}\frac{p^o_{est}}{p^o_{exp}}\right) \tag{16}$$

Fig. 3 compares the MAE and MBE for the p^o estimation methods that require T_b when the Nannoolal *et al.*[47] and the Stein and Brown[48] T_b estimation methods were used. Whiskers in Fig. 3 show the variance of compounds at the 5 % confidence interval. Therefore, if whiskers of a given p^o estimation method overlap we can be 95 % sure that there is no significant difference in its accuracy between the T_b estimation methods. The only pair of whiskers that do not overlap are for the MBE of the Myrdal and Yalkowsky[19] method, though they are very close. It was therefore concluded that there was no significant difference between the T_b estimation methods. The results presented below all use the Nannoolal *et al.*[47] T_b estimation. This method was also used to estimate T_b for the calculation of the acentric factor and critical properties in the Lee-Kesler method.

Plots of estimated p^o against measured p^o are shown in Fig. 4. Plots for Capouet and Muller[52], SIMPOL.1 and EVAPORATION have fewer points because their group contributions limited them to fewer compounds, as described in Table 2. The least-squares regression line, 1:1 line and standard deviations of estimates around the least-squares regression line are also shown in Fig. 4. Regression coefficients, standard deviations, MAE and MBE are given in Table 4. The regression coefficients (A and B) of the least-squares fit apply to the equation $\log_{10}(p^o_{est})$ = A+B($\log_{10}(p^o_{exp})$).

Of the four methods that cover the complete test set, Lee-Kesler shows the minimum MAE for the complete test set, with 0.83, while Nannoolal *et al.*[21] is the next best performing with an MAE of 0.89. With regards to MBE for 90 compounds, Nannoolal *et al.*[21] has the minimum with -0.01, while Lee-Kesler is second best with 0.03. Indeed, from Table 4 alone there is little to distinguish the performance of these two methods.

Compared to other methods Nannoolal *et al.*[21], Moller *et al.*[46] and SIMPOL show particularly poor estimation for the only two compounds with four hydrogen bonding functional groups: citric and tartaric acid (bounded by green in Fig. 4). The maximum error by any method for each compound was ranked, and plotted on the graph shown in Fig. 5, along with the errors of other methods for that compound. The estimation methods are colour coded in Fig. 5, and show a tendency for Moller *et al.*[46] to underestimate p^o, while Myrdal and Yalkowsky[19] tends to overestimate. For the compounds to which it is applicable, EVAPORATION has a notably good fit to the zero error line in Fig. 5.

Boxes numbered 1, 2 and 3 in Fig. 5 enclose compounds with a maximum error in the range, error$\geq$4, 4>error$\geq$3 and error<1, respectively, where the error is calculated by: $\log_{10}(p^o_{est})$-$\log_{10}(p^o_{exp})$ (1/Pa). Therefore, boxes 1, 2 and 3 enclose compounds with relatively very poor estimation, poor estimation and relatively good estimation, respectively. To determine what chemical property influences the accuracy of its estimation, the number of groups with hydrogen bonds per compound (Hb_n), which has been observed to affect the Nannoolal *et al.*[21] and Moller *et al.*[46] estimation accuracy previously[51], and the number of carbon atoms (C_n) per compound were plotted in the same order as maximum error ranking. A metric ($\hat{u}$) accounting for both these properties was also calculated and plotted in this order, using the equation:

$$\hat{u} = \left(\frac{Hb_n}{max(Hb_n)} + \frac{C_n}{max(C_n)} \right) /2 \tag{17}$$

where max represents the maximum number out of all compounds. Thus this equation normalises each compound's Hb_n and C_n by the maximum of each value in the test set and sums the results. The effect of the final division (/2) is to restrict results to the convenient range of 0 to 1; if a compound had both the highest number of hydrogen bonding groups and carbon atoms, its $\hat{u}$ value would be 1.

This combined metric showed a substantially better correlation with the maximum MAE ranking order (R^2=0.4, Fig. 1 of supplementary material) than the number of hydrogen bonds (R^2=0.2) and number of carbon atoms (R^2=0.0) alone, indicating that both factors combined affect estimation accuracy more strongly than separately. It is noted that increased hydrogen bonding and number of carbon atoms generally act to decrease vapour pressure. As discussed, lower vapour pressure compounds are challenging to measure, leading to measurement error and relative scarcity of measurements. Both these factors will lower the ability of estimation methods, trained on measurements, to estimate accurately. The source of inaccuracy is likely to be both errors in training set measurements (or even lack of comparable compounds in the training set) and in the functional form of estimation methods, i.e. how intermolecular forces are accounted for. The inaccuracies introduced by each will likely increase with lower volatility compounds, such that the observed trend of increased p^o inaccuracy with increased number of hydrogen bonds and carbon atoms is expected.

Seven compounds lie inside box 1 of Fig. 5, six of these are aliphatic carboxylic acids, of which five are substituted dicarboxylic acids and one is citric acid (a tricarboxylic acid). The other compound is nitrocatechol, and is easily distinguished from the others, as it tends to have an overestimated p^o, compared to the other compounds, which are mostly underestimated. Therefore, it is recommended that estimation methods focus on improving their performance for dicarboxylic acids and their oxygenated products and that more measurements of similar compounds are made to improve future assessments. It is noted, however, that p^o measurements of such compounds can be widely divergent[15], which, until resolved, will limit the extent to which estimation methods can improve through refitting.

Box 2 of Fig. 5 contains eight compounds with comparatively poorly estimated vapour pressures, and for which similar compounds should also therefore be targeted for further measurement and estimation method refitting. Of the eight compounds six are carboxylic acids (4 dicarboxylic and 2 long chain monocarboxylic). Dibutyl phthalate is also included, which is the worst estimated compound without a hydrogen bonding group. Other aromatic compounds are estimated relatively well, indicating this property is not the cause for dibutyl

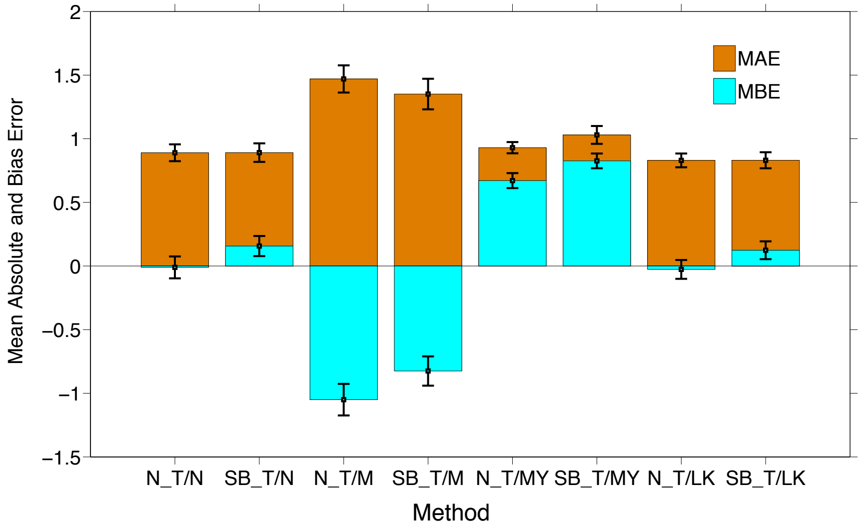

Fig. 3 The mean absolute (red) and mean bias (blue) error. Method codes have the normal boiling point method first followed by the vapour pressure estimation method, where: N_T = Nannoolal *et al.*[47], SB_T = Stein and Brown[48], and for p^o methods: N)=Nannoolal *et al.*[21], M)=Moller *et al.*[46], MY)=Myrdal and Yalkowsky[19] and LK) = Lee-Kesler. Whiskers show 5% confidence intervals.

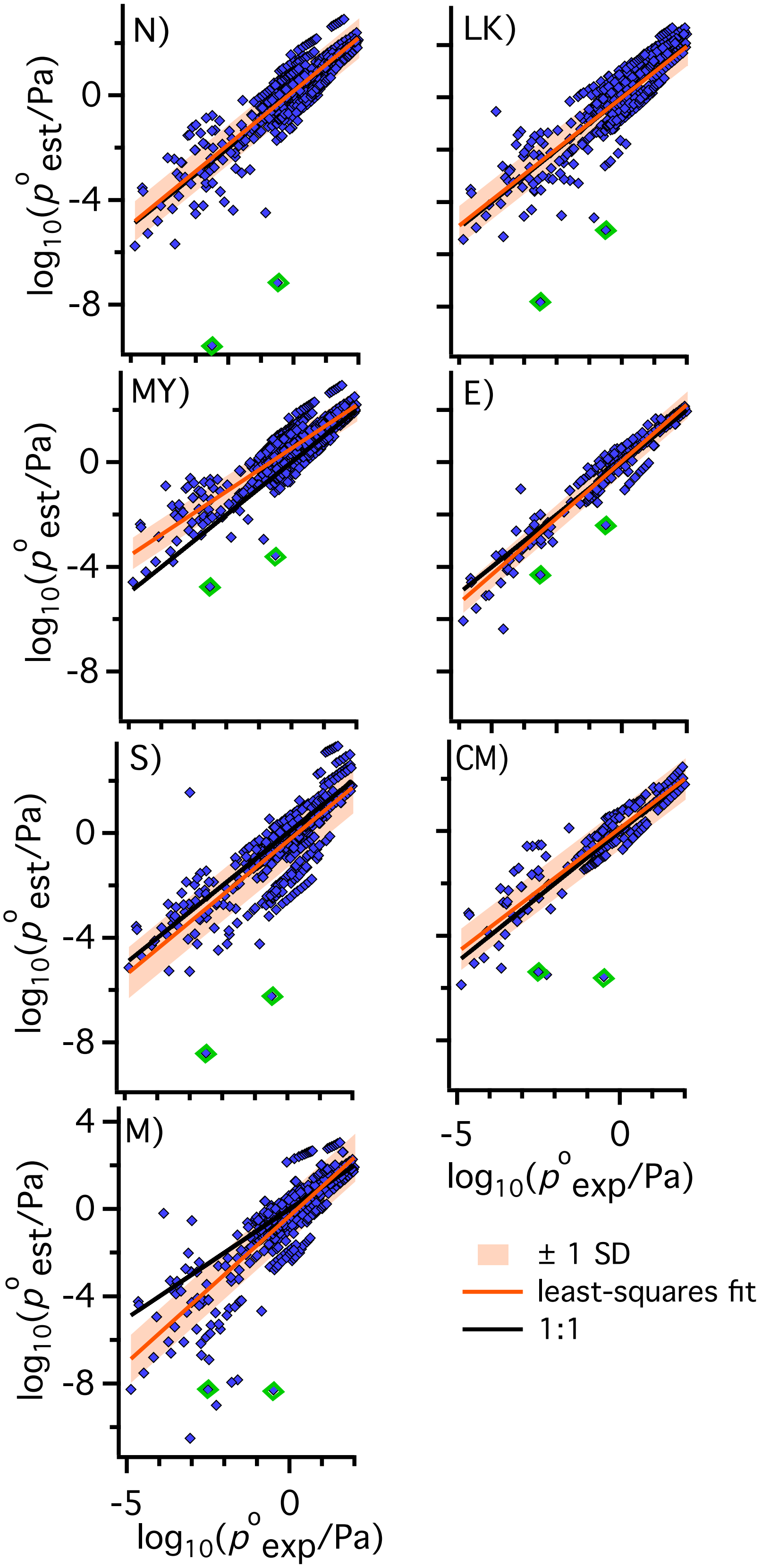

Fig. 4 Estimated vapour pressures against measured vapour pressures, where N)=Nannoolal *et al.*[21], LK)=Lee-Kesler, MY)=Myrdal and Yalkowsky[19], E)=EVAPORATION, S)=SIMPOL, CM)=Capouet and Muller[52] and M)=Moller *et al.*[46]. Orange lines are least-squares linear regression lines, shaded regions are ±1 standard deviation (SD), black lines are 1:1 fits. All abscissa ranges are consistent across plots and all ordinate ranges, except for the Moller *et al.*[46] plot, are consistent. Green bounded markers represent citric and tartaric acid.

Table 4 Performance indicators of vapour pressure estimation methods for all test set sizes: Method acronyms are the same as in Fig. 3. Statistics are: regression coefficients for the least square fit of $\log_{10}$(estimated p^o) vs. $\log_{10}$(measured p^o (x)); one standard deviation of $\log_{10}$(estimated p^o) from the least-squares regression line, the mean absolute error (MAE) and mean bias error (MBE) of estimates from measurements.

Number of Compounds	Estimation Method	Regression Coefficient	σ	MAE	MBE
90	N	0.16+1.02x	0.76	0.89	-0.01
	LK	0.15+1.04x	0.75	0.83	-0.03
	MY	0.53+0.83x	0.60	0.93	0.67
	M	-0.32+1.34x	1.10	1.46	-1.06
81	N	0.16+1.04x	0.80	0.93	0.01
	LK	0.15+1.05x	0.79	0.85	0.00
	MY	0.54+0.83x	0.64	0.98	0.74
	M	-0.33+1.38x	1.17	1.56	-1.15
	S	-0.32+1.03x	0.98	0.87	-0.29
46	N	0.37+1.04x	0.93	1.03	0.08
	LK	0.54+1.07x	0.82	0.90	0.90
	MY	0.77+0.85x	0.65	1.03	1.03
	M	-0.41+1.32x	1.41	1.90	-1.57
	S	-0.04+0.97x	0.84	0.72	0.72
	E	0.04+1.09x	0.51	0.54	0.50
42	N	0.38+1.06x	0.99	1.05	0.05
	LK	0.58+1.11x	0.85	0.84	0.12
	MY	0.76+0.86x	0.67	1.12	0.90
	M	-0.62+1.32x	1.41	1.90	-1.76
	S	-0.01+0.99x	0.89	0.73	-0.17
	E	0.07+1.11x	0.51	0.55	-0.25
	CM	0.10+0.95x	0.79	0.91	-0.15

phthalate's poor estimation. It may be the combination of an aromatic ring with comparatively long branches in dibutyl phthalate that causes poor estimation. Levoglucosan is the final compound in box 2. It has a far higher maximum error than another compound with three hydroxyl groups (1,2,3-trihydroxypropane with a maximum error of 1.27), therefore its combination of hydroxyl groups with a bicyclic structure and/or ether groups may cause its high inaccuracy.

The 27 compounds in box 3 of Fig. 5 have their vapour pressures estimated relatively accurately by all methods. These compounds are relatively simple, with an average number of hydrogen bonding groups of 0.9 compared to the entire test set average of 1.5. Six compounds have 2 hydrogen bonding groups (the maximum number for box 3 compounds), and of these five are dihydroxyl compounds. Although there are four monocarboxylic compounds in box 1 there are no polycarboxylics. Thus, the estimation methods perform much better for dihydroxyl compounds than dicarboxylic ones. Despite the low number of hydrogen bonding groups, all but one of the 27 compounds has multiple oxygenating groups, including ketones, amines and halogen groups. 2-phenylbromide-TEGMME, for example, contains four ether groups and a bromine group. The relatively good estimation of these compounds is further evidence that estimation methods perform comparatively poorly for compounds with increased number of hydrogen bonding groups, and therefore decreased volatility.

The results of the assessment of estimates against measurements for compounds with $p^o_{298.15K} < 10^{-2}$ Pa only are given in Table 5. Of the methods that cover all 50 of these compounds, Lee-Kesler again has the lowest MAE, and Nannoolal *et al.*[21] the lowest MBE. Focusing on these lower volatility compounds increases the MAE for Nannoolal *et al.*[21] and Myrdal and Yalkowsky[19]. In contrast, the MAE for Lee-Kesler is constant between the low volatility subset and the complete test set. Meanwhile, the positive bias of Myrdal and Yalkowsky[19] increases substantially from the complete test set to the low volatility subset, which can also be seen in Fig. 4. However, for both test sets, it shows a notably lower standard deviation around its least square regression line than the other methods.

Table 5 Performance indicators for estimation methods applied to compounds with $p^o_{298.15K} < 10^{-2}$ Pa. See Table 4 and Fig. 4 for the definitions of indicators and method acronyms, respectively. Results are first given for the maximum number of low volatility compounds estimation methods were applicable to. Below these results are given for the subset of 33 compounds that the Capouet and Muller[52] method was applicable to.

Number of compounds	Estimation Method	Regression Coefficients	σ	MAE	MBE
50	N	0.54+1.14x	0.84	0.97	0.03
50	LK	0.38+1.08x	0.69	0.82	0.08
50	MY	0.94+0.94x	0.58	1.06	0.95
50	M	-0.31+1.36x	1.41	1.67	-1.16
49	S	-0.10+0.97x	0.76	0.78	-0.19
34	E	-0.07+1.08x	0.53	0.60	-0.35
33	CM	0.00+0.91x	0.81	0.77	-0.07
33	N	0.61+1.15x	1.00	1.04	-0.13
33	LK	0.52+1.10x	0.75	0.79	-0.02
33	MY	1.03+0.95x	0.65	1.05	0.88
33	M	-0.93+1.19x	1.46	1.94	-1.80
33	S	0.09+1.02x	0.80	0.71	-0.27
33	E	-0.07+1.09x	0.53	0.61	-0.37

For the combined methods inaccuracies result from either the vapour pressure or the normal boiling point equation. There is a scarcity of T_b estimation method assessments relevant to SVOC, indeed the authors are not aware of any. This may be due to measurement difficulties, as these compounds have a tendency to decompose at temperatures approaching their boiling point. Here we briefly assessed T_b estimates by collating normal boiling points from the online chemistry

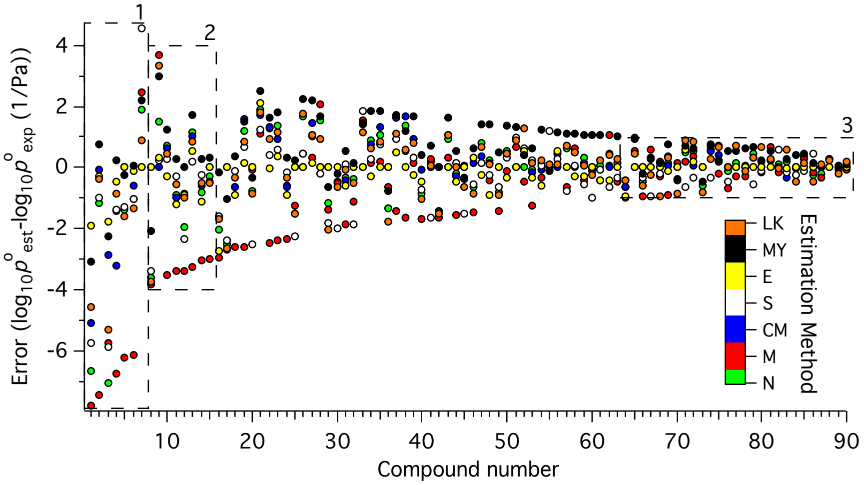

Fig. 5 The vapour pressure error of each compound for each estimation method (averaged where necessary over multiple measurement temperatures). Estimation methods are colour coded, as shown in the legend. Boxes segregate compounds falling into error ranges as discussed in the main text. Compounds have been ordered according to their maximum error.

databases of ChemSpider[59] and NIST[60]. Of the 90 compounds in our vapour pressure test set 30 had experimental T_b values in at least one database. Similar to vapour pressures, the primary source can be difficult to determine, and the claim of an experimental value difficult to verify as measurements are frequently unaccompanied by error margins. The NIST database reports the range of values around an average if several sources are available for a T_b value. The largest range was ± 70 K for triacetin, however, a more typical range around single values (where available) was ± 3 K.

The Stein and Brown[48] and Nannoolal *et al.*[21] T_b estimates for the 30 compounds had the experimental values substituted and the resulting residuals were plotted against measured compound p_o in Fig. 6. If a compound had measurements at multiple temperatures, the p_o at the median temperature was taken, so that the plot was most easily comparable to those in Fig. 4. The majority of residuals are within ± 50 K, but 2-methyl succinic acid has a value of ∼180 K for both estimates. A similar compound, succinic acid, which we expect to have a T_b within 20 K of 2-methyl succinc acid (based on experience and estimates), has a database T_b of 508 K, compared to 380 K for the latter. The residual for succinic acid is within the typical range for both estimates. Thus the 2-methyl succinic acid appears to be due to measurement error, or derived erroneously by estimation. Consequently, the residual for 2-methyl succinic acid was omitted from calculations of average absolute errors; these were 19 K and 13 K for Stein and Brown[48] and Nannoolal *et al.*[47], respectively.

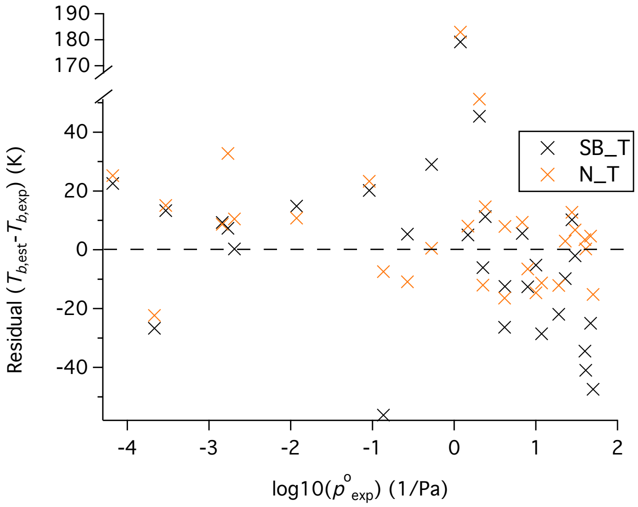

Fig. 6 The residual of Stein and Brown[48] (black crosses) and Nannoolal *et al.*[47] (orange crosses) from measured normal boiling points against the vapour pressures of the corresponding compounds. The dotted line represents zero disagreement between estimate and measurement.

The experimental T_b values were input to the combined p_o estimation methods and the resulting errors from experimental p_o values were found. No method showed an improvement in p_o accuracy for all compounds using experimental T_b. The error using experimental T_b was divided by that using estimated T_b values, so that a value of 1 represented no change and a value of 0.5, for example, represented a halving of error. The median values for Nannoolal *et al.*[21] and Myrdal and Yalkowsky[19] were 1.3 and 1.4, respectively, and those for Moller *et al.*[46] and Lee-Kesler were 0.7 and 0.8, respectively. For the majority of the 30 compounds, the former two methods, therefore, had a lower error using estimated T_b than experimental, whilst for the latter two methods p_o accuracy improved for the majority of compounds when experimental T_b was used.

Decoupling the contribution of error from the T_b estimation and the p_o equation is not straightforward since, as the previous paragraph shows, inaccuracy in estimated T_b can actually improve estimated p_o accuracy if it offsets the inaccuracy of the p_o equation. In effect, the contribution of inaccuracy from the T_b estimation is a function of the p_o estimation method and the compound in question, since it depends on how the method treats a given compound. The relative contributions of error will therefore vary by p_o estimation method and compound.

Were one of the combined methods to be used for p_o estimation in a chemical transport model using a mechanistic approach to organic aerosol partitioning the T_b of SVOCs would also have to be estimated. Therefore the results of this study are pertinent to such models.

Moller *et al.*[46] vapour pressure results show a notably larger standard deviation than any other method. A large deviation of Moller *et al.*[46] p^o estimates from Capouet and Muller[52] estimates was observed in Compernolle *et al.*[45], in addition to unrealistically high estimates for molecules with more than three oxygenated functional groups. In our assessment major errors arise from underestimation, with a tendency to increasingly underestimate p^o with decreasing measured p^o. Generally, lower vapour pressures occur at temperatures comparatively far from the compound's normal boiling point. Under this situation error introduced by either (or both of) the estimation of T_b or the p^o curve is likely to be magnified since the Moller *et al.*[46] method works by extrapolating the p^o vs T curve from the normal boiling point.

Myrdal and Yalkowsky[19] has a lower standard deviation than Nannoolal *et al.*[21] and Lee-Kesler, however, it shows an increasing tendency to overestimate with decreasing p^o, as also observed in Compernolle *et al.*[45] and Barley and McFiggans[10]. Compernolle *et al.*[45] suggest this could be due to the assumption of a constant change in heat capacity from gas to liquid with temperature. Since the tendency is also observed in the current assessment we recommend this assumption be tested using measurement data. Myrdal and Yalkowsky[19] estimates for citric and tartaric acid p^o show the best agreement with measurements (of the methods covering the complete test set), suggesting it could produce relatively accurate estimates for compounds with high numbers of hydrogen bonding groups. Fig. 5 supports this advantageous property of Myrdal and Yalkowsky[19], since this method shows relatively low errors for compounds in the high maximum error boxes (box numbers 1 and 2). It seems likely, however, that this relatively good accuracy is partly attributable to an offset by its systematic overestimation of p^o.Fig. 6 shows no systematic bias in T_b estimates with compound vapour pressure, indicating that the main cause of the overestimation in Myrdal and Yalkowsky[19] is the p^o equation.

Based on the above results, of the methods assessed here that cover the complete test set, the Lee-Kesler or Nannoolal *et al.*[21] method is recommended. Compared to Nannoolal *et al.*[21], Lee-Kesler shows very similar mean absolute error (0.83:0.89), standard deviation (0.75:0.76) and regression coefficients values (0.15+1.04x:0.16+1.02x) for 90 compounds (Table 4). There is is more distinction between the performances of these methods when subsets of the data are analysed. For the low volatility subset (Table5), the Nannoolal *et al.*[21] MAE rises to 0.97, while Lee-Kesler changes little from the complete test set, with 0.82. Nannoolal *et al.*[21], however, still outperforms Lee-Kesler on MBE (0.03:0.08). Regarding Fig. 5 and the compounds in boxes 1 and 2, which have the highest maximum estimation error and tend to have the highest number of hydrogen bonding groups and/or carbon atoms, Lee-Kesler generally performs better than Nannoolal *et al.*[21]. Fig. 4 shows for the two compounds with 4 hydrogen bonding groups it has a lower MAE by about 2 error units. Therefore, based on these results, Lee-Kesler is recommended over Nannoolal *et al.*[21].

This contrasts with the conclusion of Barley and McFiggans[10], which recommended Nannoolal *et al.*[21] over Lee-Kesler. The reason for this change is the expanded test set. As shown in Fig. 1, the main difference between this test set and 2010 study is the decrease in mean average compound vapour pressure. Table 5 and Fig. 4 show that Lee-Kesler produces more accurate results for these lower volatility compounds, most likely because it accounts for the effect of their increased oxygenation and/or chain length better. Therefore it is the inclusion of lower volatility compounds that has favoured Lee-Kesler in this study.

Our recommendation is coupled with a request for further assessment following further vapour pressure measurements. As discussed, the full test set is biased toward carboxylic acids, and is without some functional groups known to be present in the atmosphere (such as hydroperoxide[61]). The low volatility test set is particularly skewed toward methods that accurately estimate for carboxylic acids, with 42 of the 50 compounds containing a carboxylic acid group. Thus, further measurements, particularly for non-carboxylic compounds, are required for more rigorous future assessment.

Were the systematic overestimation by the Myrdal and Yalkowsky[19] extrapolation method to be rectified this would likely be recommended due to its relatively low standard deviation: approximately 20 % lower than that of Lee-Kesler and Nannoolal *et al.*[21] for the 90 compound test set, and the same improvement over Lee-Kesler for the 50 low volatility subset also.

EVAPORATION shows clear superiority of all methods in standard deviation and MAE for the test sets with 46 and 42 compounds. For 46 compounds the intercept for its least-squares fit line is equal with SIMPOL.1 in being the closest to 0. Its gradient is within 0.05 and 0.02 of Nannoolal *et al.*[21] and Lee-Kesler, respectively, therefore showing a comparatively good fit to the 1:1 line. For the low volatility subset, EVAPORATION again shows superiority in MAE when compared against other methods for the same set of compounds (Table 5). It is notable, however, that its bias is comparatively low for this subset, and its regression coefficients in both Tables 4 and5 and its least square line in Fig. 3, indicate a tendency to increasingly underestimate vapour pressure with decreasing vapour pressure. Nevertheless, these results show that over all compounds to which it is applicable EVAPORATION provides the most accurate p^o values of the estimation methods assessed here. It is recommended that this method be extended to cover all functionalities found in SVOC since its good accuracy here suggests potential to estimate p^o with improved accuracy for all SVOC. Additional work may be required, however, to rectify its tendency to underestimate vapour pressure for lower volatility compounds.

As discussed in section 2.1, the EVAPORATION method was produced using some of the measurements it is assessed against. This is likely to give it an advantage over methods produced earlier, such as Lee-Kesler, particularly for carboxylic acids, for which there has been a concerted measurement effort recently. Therefore, the comparatively high accuracy of EVAPORATION is, to an extent, artificial. Consequently, it is strongly recommended to test EVAPORATION against further measurements, particularly for compounds other than carboxylic acids, but relevant to the atmosphere.

A further consideration is reliability in the measured p^o. Although we have assumed zero error in measurements since our objective was to assess estimates, uncertainties in measurements can span orders of magnitude. There is a wide range in reported values for many compounds (for example 4 orders of magnitude for azelaic acid between Salo *et al.*[62] and Cappa *et al.*[63]). Experimental issues arise from the following: extrapolation to 298 K from higher temperatures; unknown or poorly controlled sample state (e.g. amorphous or hydrated material); contamination of sample with higher volatility material; contamination or saturation of the instrument; availability of low vapour pressure calibration compounds. The measurement of low volatility compounds remains an active area of research and is the subject of a recent review paper[38].

3.2 Sensitivity of Particulate Mass Loading and Composition to Estimation Methods.

Measured and estimated p^o for the test set compounds were input to a molar-based absorptive partitioning model[55] as described in section 2.3. Results were converted to mass loadings to allowing comparison of simulated particulate concentrations. Compounds were fitted with concentrations based on the organic aerosol concentration distribution with volatility reported in the ambient field study of Cappa and Jimenez[58]. The results for a low (remote scenario) and high (polluted scenario) particulate loading case at 60 % relative humidity are shown in Fig. 7. When using measured p^o values, the low and high cases had secondary organic aerosol mass loadings (SOA) of 1.5 μg m^{-3} and 48 μg m^{-3} respectively, and the SOA:non-volatile mass ratio was 1:1.

Table 6 presents the fraction change in SOA mass loading relative to the measured p^o case for the different methods.

Fig. 7 and Table 6 show that the performance of estimation methods to predict pure compound vapour pressures is not linearly related to their performance to predict SOA loading when their estimated vapour pressures are input to an absorptive partitioning model. Most estimation methods overestimate SOA mass loading. This is true even for methods giving a net positive bias in Table 4, and which one might therefore expect to underestimate SOA. The cause is the fitting of compound concentrations to field observations of increasing concentration with increasing p^o. To exemplify the effect of this, consider the case where a negligible bias is produced through the same number of compounds having vapour pressure underestimated as overestimated by the same amount. The underestimated compounds generate more condensed mass than the overestimated compounds they have replaced in p^o space would have done (with measured p^o) since they have the greater total concentration. Consequently there has to be a relatively major systematic overestimation of p^o for SOA to be underestimated, as seen with Myrdal and Yalkowsky[19] in Table 6.

The variation in SOA mass loading accuracy between the complete and reduced test sets is likely due to an average change in the composition of compounds between the sets. For certain classes of compound a given method's accuracy will vary (e.g. Asher *et al.*[64]). The variation in SOA accuracy between the remote and polluted scenarios is due to a change in the average chemical composition of SVOC. This is because the vapour pressure range in which SVOC reside shifts depending on the particulate loading (C_{MOA} (eq. 13)). The issue of changing chemical composition between the reduced and complete test set is therefore an artefact of the assessment. In contrast, the issue of changing composition with volatility is expected in the atmosphere: lower volatility compounds are expected to be more oxidised and/or have higher

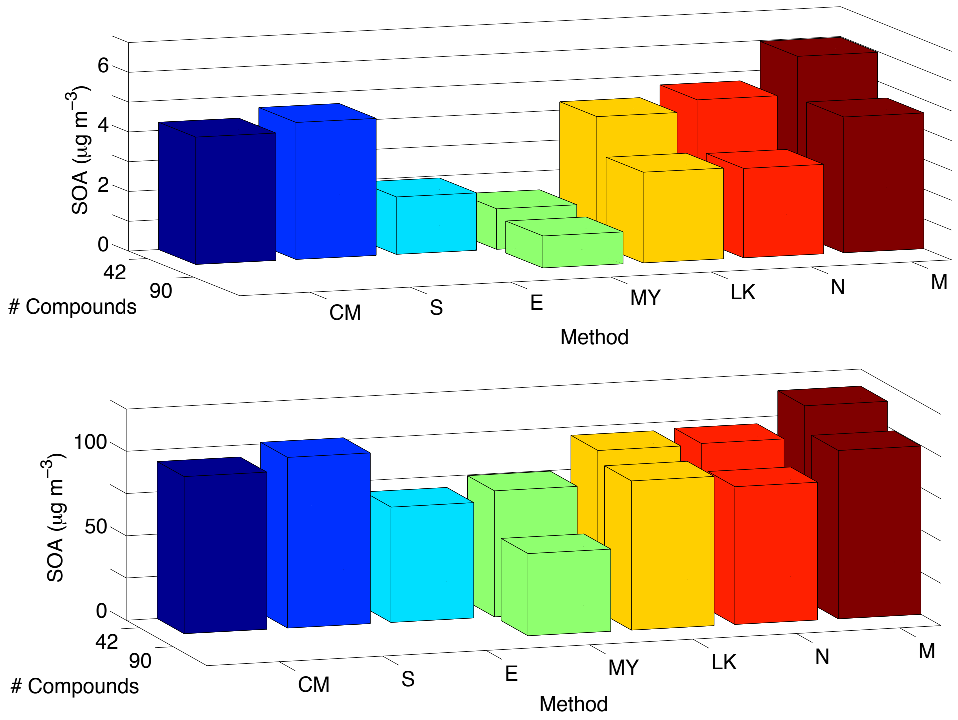

Fig. 7 Modeled secondary organic aerosol mass loading (SOA) using different vapour pressure estimation methods. Method acronyms are the same as in Fig. 4. The upper and lower plots are for compound concentrations that yielded 1.5 and 48 μgm^{-3} SOA (when measured vapour pressures were used), respectively. Both simulations were at a relative humidity of 60 % and when measured p^o were used the non-volatile:SOA mass concentration ratio was 1:1.

molecular weights.

Despite comparatively low MAE and MBE values for the vapour pressure comparisons Nannoolal *et al.*[21] and Lee-Kesler show relatively poor accuracy for SOA mass loading. This is because of the increasing total mass of compounds with increasing volatility (as discussed above), combined with the lack of any compensating effect such as a positive p^o bias. They show an improved SOA accuracy for the polluted case, indicating a decreasing MAE for compounds with higher volatilities since these contribute more to the condensed phase under a higher particulate loading. Fig. 4 demonstrates this is indeed the case.

Myrdal and Yalkowsky[19] and EVAPORATION generate the most accurate SOA mass loading estimates. The positive p^o bias of the former causes an underestimation of the SOA in the low loading scenario. In the high loading scenario the upward shift of the volatility range of SVOC results in a lower average overestimation of SVOC p^o by Myrdal and Yalkowsky[19] (Fig. 4). This leads to a very accurate loading estimation for the complete test set and an overestimation of loading in the reduced test set (Table 6).

Despite EVAPORATION having a negative MBE for the reduced test set (Table 4), the relatively low MAE of this method means that the movement of compounds in volatility space between the measured and estimated p^o case is relatively minimal.

The composition and not only mass of the modeled SOA is also important. To characterise composition, the average carbon oxidation state and number of carbon atoms of compounds in the condensed phase was found and is shown in Fig. 8. The average is weighted by the condensed concentration of each compound. The first feature to note in Fig. 8 is that markers of the same shape (marker shape represents scenarios) generally congregate together. The clustering of the 90 compound results vs. those of the 42 compounds is, like the SOA loading accuracy results, an artefact of the test set (i.e., due to the sample of compounds in each set). Secondly, the estimation methods show a strong tendency to overestimate the mean carbon oxidation state and underestimate the number of carbon atoms per compound compared to the base case (in which measured vapour pressures are used). This tendency is generally reduced, to an extent, when the loading is increased, indicating less sensitivity of composition to vapour pressure estimation when partitioning coefficients increase, as for SOA loading. The direction of points along this plot approximately follows the trajectory of where biomass burning or-

Table 6 The fraction change from using the measured p^o to estimated p^o to obtain SOA mass loading, where fraction change = SOA loading using estimated p^o/SOA loading using measured p^o (e.g. 0.5 equates to a factor of 2 decrease). See Fig. 4 for method acronyms.

Measured p^o Case SOA (μgm^{-3})	Number of Compounds	Method	Fraction Change
1.5	42	N	3.1
		M	4.0
		LK	2.9
		MY	0.9
		EVA	1.3
		SIM	3.1
		CM	2.8
	90	N	2.0
		M	3.0
		LK	2.0
		MY	0.7
48.0	42	N	2.0
		M	2.4
		LK	2.0
		MY	1.6
		EVA	1.4
		SIM	2.1
		CM	1.9
	90	N	1.7
		M	2.1
		LK	1.8
		MY	1.0

ganic aerosol (BBOA) (lower average oxidation state, higher number of carbon atoms), SVOC and low-volatility oxidised organic aerosol (LV-OOA) (higher oxidation state, lower number of carbon atoms) would be expected to be found[37]. The base case points fall within the BBOA region, whereas the majority of estimation method points fall within the SVOC or even LV-OOA region. Thus, the differences between the base case and the majority of estimation methods in Fig. 8 represent major differences in the SOA composition.

There is little variation between the combined methods in Fig. 8, though, for the low loading scenarios, Myrdal and Yalkowsky[19] shows notably better agreement with the base case in number of carbon atoms and slightly better agreement in average oxidation state. This appears to reinforce its choice as the most useful method for estimating SOA loading and composition (despite it showing comparatively poor accuracy for pure compound p^o). EVAPORATION has very good agreement in both dimensions for the low loading scenario. For the high loading scenario its agreement is worse but notably better than the other methods. The utility of EVAPORATION is, however, limited by the number of functional groups it covers.

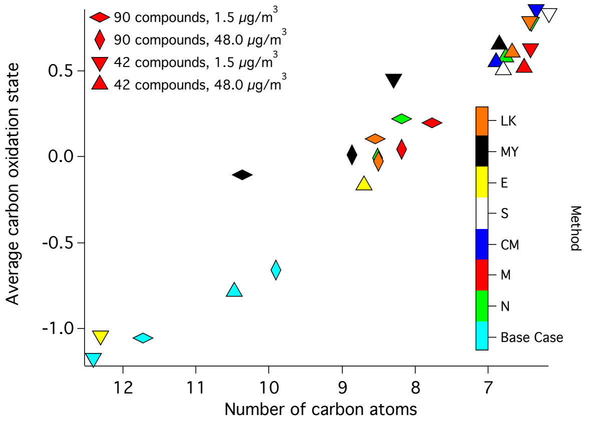

Fig. 8 The weighted average number of carbon atoms and average oxidation state in the condensed phase. Marker shapes represent different scenarios (number of compounds in the test set and loading), while marker colours represent the method used to obtain vapour pressures (as shown on the colour scale).

There has to be a clear distinction on method recommendation based on 1) pure component comparisons and 2) atmospheric SOA predictions using empirically constrained volatility profiles. Overall EVAPORATION or Myrdal and Yalkowsky[19] have been shown as the preferred methods for estimating SOA (point 2). As discussed, Myrdal and Yalkowsky[19] generates the most accurate loading through the offsetting effect of its overestimation of p^o for most compounds. Despite this, it also boasts comparatively good estimates of average oxidation state and number of carbon atoms for low loading scenarios. The Myrdal and Yalkowsky[19] method, however, requires further assessment against SVOCs with measured $p^o < 10^{-1}$ Pa. Its systematic overestimation of p^o on SOA when these compounds are included is expected to produce even greater underestimation of loading than seen here. For the other estimation methods, when applied to a test set more representative of the atmosphere, the greatest change is likely to be the introduction of more oxygenated compounds. As discussed, the methods tend to show increasing error for compounds of increasing number of hydrogen bonds, thus a greater mean absolute error of vapour pressure and overestimation of SOA loading would be expected from inclusion of more oxygenated compounds.

The above discussion demonstrates that absolute accuracy as well as bias influences the accuracy of SOA mass loading estimates. Sensitivity of predicted SOA mass, and composition, to choice of estimation method depends on the abundance and volatility of each compound. For example, estimation methods have been applied to predicted compounds generated by near-explicit models of VOC oxidation: the Master Chemical Mechanism in Barley *et al.*[49] and GECKO-A in Valorso *et al.*[56]. In the former study a factor of 10 difference was noted in SOA between Myrdal and Yalkowsky[19] and Nannoolal *et al.*[21] when both used Nannoolal *et al.*[47] T_b estimates. Valorso *et al.*[56] found the difference in SOA is greatest when the aerosol loading is low, varying by a maximum factor of 3.5 between Myrdal and Yalkowsky[19] with Joback and Reid[50] T_b and Nannoolal *et al.*[21] with Nannoolal *et al.*[47] T_b. When aerosol loading is low, compounds with relatively lower vapour pressures are able to condense whereas those with higher vapour pressures cannot. As others and we have shown, the discrepancy between estimation methods increases as vapour pressure decreases so this is expected. As Barley *et al.*[49] found, as the number of compounds increases, sensitivity to estimation method decreases.

4 Conclusion

This study compared the vapour pressures generated by seven estimation methods against measurements to assess their accuracy. Compared to the assessment of Barley and McFiggans[10] the number of test set compounds was doubled and the types and frequency of functional groups per compound was increased, thereby providing a more thorough assessment. Three of the methods were not applicable to all the functional groups present in the test set of compounds, meaning that a reduced test set was also needed. EVAPORATION gave the minimum MAE for the reduced test set, while Lee-Kesler gave the minimum MAE for the complete set. The MBE was lowest for Nannoolal *et al.*[21], however this method showed comparatively poor accuracy for compounds with lower volatility. Similarly, the Myrdal and Yalkowsky[19] method increasingly overestimated vapour pressures for compounds with decreasing volatility.

Using a simple absorptive equilibrium model, the sensitivity of predicted SOA loading to choice of estimation technique was assessed by mapping concentrations of compounds within the test set to a volatility profile derived from an ambient study. By comparing predictive saturation vapour pressures with measured vapour pressures, the majority of methods, including Lee-Kesler and Nannoolal *et al.*[21], showed overestimation of SOA by factors of between 1.8 - 3.

Myrdal and Yalkowsky[19] and EVAPORATION showed the highest accuracy for SOA loading estimates using our methodology as well as the highest accuracy for SOA composition (represented by the average carbon oxidation state and number of carbon atoms per compound). The effect of the increasing overestimation of vapour pressure with decreasing volatility by Myrdal and Yalkowsky[19] on SOA loading, however, requires testing, since none of the compounds here had volatilities toward the lower end of the SVOC volatility region. The EVAPORATION method cannot be used on compounds containing certain functional groups. If it could be extended to cover all groups present in the atmosphere and maintain its comparatively high accuracy it would be the preferable method for both vapour pressure estimation and SOA modeling.

The recommended method is dependent on its application: for a single organic compound EVAPORATION has been shown as most accurate and consistent. However, if it is not applicable to the compound in question due to missing functional group information, Lee-Kesler is recommended. On the contrary, based on SOA mass loading sensitivities, Myrdal and Yalkowsky[19] has been shown as best. A future study should consider whether the Myrdal and Yalkowsky[19] method coupled with a chemical degradation model or the volatility basis set provides the better estimation of mass distribution by volatility in a chemical transport model.

Acknowledgments

The Natural Environment Research Council funded this work through the studentship of Simon O'Meara and grant numbers: NE/J02175X/1, NE/J009202/1 and NE/H002588/1.

References

1 P. Forster and V. Ramaswamy and P. Artaxo and T. Berntsen and R. Betts and D.W. Fahey and J. Haywood and J. Lean and D.C. Lowe and G. Myhre and J. Nganga and R. Prinn and G. Raga and M. Schulz and R. Van Dorland, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007, ch. 2, pp. 129–234.

2 C. Pope and D. Dockery, *Journal of the Air & Waste Management Association*, 2006, **56**, 709–742.

3 M. R. Heal, P. Kumar and R. M. Harrison, *Chemical Society reviews*, 2012, **41**, 6606–30.

4 M. Andreae and P. J. Crutzen, *Science*, 1997, **276**, 1052–1058.

5 S. Fuzzi, M. Andreae, B. Huebert, M. Kulmala, T. Bond, M. Boy, S. Doherty, A. Guenther, M. Kanakidou, K. Kawamura, V. Kerminen, U. Lohmann, L. Russell and U. Poschl, *Atmospheric Chemistry and Physics*, 2006, **6**, 2017–2038.

6 B. Aumont, S. Szopa and S. Madronich, *Atmospheric Chemistry and Physics*, 2005, **5**, 2497–2517.

7 Q. Zhang, J. L. Jimenez, M. R. Canagaratna, J. D. Allan, H. Coe, I. Ulbrich, M. R. Alfarra, a. Takami, a. M. Middlebrook, Y. L. Sun, K. Dzepina, E. Dunlea, K. Docherty, P. F. DeCarlo, D. Salcedo, T. Onasch, J. T. Jayne, T. Miyoshi, a. Shimono, S. Hatakeyama, N. Takegawa, Y. Kondo, J. Schneider, F. Drewnick, S. Borrmann, S. Weimer, K. Demerjian, P. Williams, K. Bower, R. Bahreini, L. Cottrell, R. J. Griffin, J. Rautiainen, J. Y. Sun, Y. M. Zhang and D. R. Worsnop, *Geophysical Research Letters*, 2007, **34**, L13801.

8 M. Hallquist, J. C. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, J. Dommen, N. M. Donahue, C. George, A. Goldstein, J. Hamilton, H. Herrmann, T. Hoffmann, Y. Iinuma, M. Jang, M. Jenkin, J. Jimenez, A. Kiendler-Scharr, W. Maenhaut, G. McFiggans, T. F. Mentel, A. Monod, A. Prevot, J. Seinfeld, J. Surratt, R. Szmigielski and J. Wildt, *Atmospheric Chemistry and Physics*, 2009, 5155–5236.

9 N. M. Donahue, a. L. Robinson, C. O. Stanier and S. N. Pandis, *Environmental Science & Technology*, 2006, **40**, 2635–43.

10 M. H. Barley and G. McFiggans, *Atmospheric Chemistry and Physics*, 2010, **10**, 749–767.

11 M. Barley, D. O. Topping, M. E. Jenkin and G. McFiggans, *Atmospheric Chemistry and Physics*, 2009, **9**, 2919–2932.

12 A. Baklanov, K. . Schlünzen, P. Suppan, J. Baldasano, D. Brunner, S. Aksoyoglu, G. Carmichael, J. Douros, J. Flemming, R. Forkel, S. Galmarini, M. Gauss, G. Grell, M. Hirtl, S. Joffre, O. Jorba, E. Kaas, M. Kaasik, G. Kallos, X. Kong, U. Korsholm, a. Kurganskiy, J. Kushta, U. Lohmann, a. Mahura, a. Manders-Groot, a. Maurizi, N. Moussiopoulos, S. T. Rao, N. Savage, C. Seigneur, R. S. Sokhi, E. Solazzo, S. Solomos, B. Sø rensen, G. Tsegas, E. Vignati, B. Vogel and Y. Zhang, *Atmospheric Chemistry and Physics*, 2014, **14**, 317–398.

13 G. McFiggans, D. O. Topping and M. H. Barley, *Atmospheric Chemistry and Physics*, 2010, **10**, 10255–10272.

14 A. Delle Site, *Journal of Physical and Chemical Reference Data*, 1997, **26**, 157.

15 A. J. Huisman, U. K. Krieger, a. Zuend, C. Marcolli and T. Peter, *Atmospheric Chemistry and Physics Discussions*, 2013, **13**, 1133–1177.

16 M. Camredon and B. Aumont, *Atmos. Environ.*, 2006, **40**, 2105–2116.

17 A. K. Mbienda, C. Tchawoua, D. Vondou and F. M. Kamga, *International Journal of Geophysics*, 2013, **2013(612375)**, 13.

18 R. Reid, J. Prausnitz and B. Poling, *The Properties of Gases and Liquids*, McGraw-Hill, New York and London, 4th edn, 1987.

19 P. Myrdal and S. Yalkowsky, *Industrial Engineering Chemical Research*, 1997, **36**, 2494–2499.

20 S. Compernolle, K. Ceulemans and J.-F. Müller, *Atmospheric Chemistry and Physics*, 2011, **11**, 9431–9450.

21 Y. Nannoolal, J. Rarey and D. Ramjugernath, *Fluid Phase Equilibria*, 2008, **269**, 117–133.

22 B. J. Murray and A. K. Bertram, *Physical Chemistry Chemical Physics : PCCP*, 2008, **10**, 3287–301.

23 A. Virtanen, J. Joutsensaari, T. Koop, J. Kannosto, P. Yli-Pirilä, J. Leskinen, J. M. Mäkelä, J. K. Holopainen, U. Pöschl, M. Kulmala, D. R. Worsnop and A. Laaksonen, *Nature*, 2010, **467**, 824–7.

24 C. Marcolli, B. Luo and T. Peter, *The Journal of Physical Chemistry A*, 2004, **108**, 2216–2224.

25 C. D. Cappa, E. R. Lovejoy and A. R. Ravishankara, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 18687–18691.

26 J. Prausnitz, R. Lichtenhaler and E. de Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edn, 1986.

27 N. L. Ng, M. R. Canagaratna, Q. Zhang, J. L. Jimenez, J. Tian, I. M. Ulbrich, J. H. Kroll, K. S. Docherty, P. S. Chhabra, R. Bahreini, S. M. Murphy, J. H. Seinfeld, L. Hildebrandt, N. M. Donahue, P. F. DeCarlo, V. a. Lanz, a. S. H. Prévôt, E. Dinar, Y. Rudich and D. R. Worsnop, *Atmospheric Chemistry and Physics*, 2010, **10**, 4625–4641.

28 N. M. Donahue, S. a. Epstein, S. N. Pandis and a. L. Robinson, *Atmospheric Chemistry and Physics*, 2011, **11**, 3303–3318.

29 M. C. Pietrogrande, D. Bacco, M. Visentin, S. Ferrari and V. Poluzzi, *Atmospheric Environment*, 2014, **86**, 164–175.

30 S. Decesari, S. Fuzzi, M. Facchini, M. Mircea, L. Emblico, F. Cavalli, W. Maenhaut, X. Chi, G. Schkolnik, A. Falkovich, Y. Rudich, M. Claeys, V. Pashynska, G. Vas, I. Kourtchev, R. Vermeylen, A. Hoffer, M. Andreae, E. Tagliavini, F. Moretti and P. Artaxo, *Atmospheric Chemistry and Physics*, 2006, **6**, 375–402.

31 D. C. S. Beddows, R. J. Donovan, R. M. Harrison, M. R. Heal, R. P. Kinnersley, M. D. King, D. H. Nicholson and K. C. Thompson, *Journal of Environmental Monitoring*, 2004, **6**, 124–33.

32 J. Hamilton, P. Webb, a. Lewis, J. Hopkins, S. Smith and P. Davy, *Atmospheric Chemistry and Physics Discussions*, 2004, **4**, 1393–1423.

33 W. F. Rogge, M. A. Mazurek and L. M. Hildemannt, *Atmospheric Environment*, 1993, **27A**, 1309–1330.

34 D. Helmig, A. Bauer, J. Muller and W. Klein, *Atmospheric Environment*, 1990, **24**, 179–184.

35 M. Kalberer, D. Paulsen, M. Sax, M. Steinbacher, J. Dommen, a. S. H. Prevot, R. Fisseha, E. Weingartner, V. Frankevich, R. Zenobi and U. Baltensperger, *Science*, 2004, **303**, 1659–62.

36 D. a. Day, S. Liu, L. M. Russell and P. J. Ziemann, *Atmospheric Environment*, 2010, **44**, 1970–1979.

37 J. H. Kroll, N. M. Donahue, J. L. Jimenez, S. H. Kessler, M. R. Canagaratna, K. R. Wilson, K. E. Altieri, L. R. Mazzoleni, A. S. Wozniak, H. Bluhm, E. R. Mysak, J. D. Smith, C. E. Kolb and D. R. Worsnop, *Nature chemistry*, 2011, **3**, 133–9.

38 M. Bilde, *To be submitted*.

39 A. Kettle and M. Andrea, *Journal of Geophysical Research*, 2000, **105**, 26793–26808.

40 K. Barsanti, P. McMurry and J. Smith, *Atmospheric Chemistry and Physics*, 2009, **9**, 2949–2957.

41 M. C. Facchini, S. Decesari, M. Rinaldi, C. Carbone, E. Finessi, M. Mircea, S. Fuzzi, F. Moretti, E. Tagliavini, D. Ceburnis and C. D. O'Dowd, *Environmental Science & Technology*, 2008, **42**, 9116–21.

42 J. D. Surratt, Y. Gómez-González, A. W. H. Chan, R. Vermeylen, M. Shahgholi, T. E. Kleindienst, E. O. Edney, J. H. Offenberg, M. Lewandowski, M. Jaoui, W. Maenhaut, M. Claeys, R. C. Flagan and J. H. Seinfeld, *The Journal of Physical Chemistry A*, 2008, **112**, 8345–78.

43 S. Decesari, M. Mircea, F. Cavalli, S. Fuzzi, F. Moretti, E. Tagliavini and M. C. Facchini, *Environmental Science & Technology*, 2007, **41**, 2479–84.

44 P. Saxena and L. M. Hildemann, *Journal of Atmospheric Chemistry*, 1996, **24**, 57–109.

45 S. Compernolle, K. Ceulemans and J.-F. Müller, *Atmospheric Chemistry and Physics*, 2010, **10**, 6271–6282.

46 B. Moller, J. Rarey and D. Ramjugernath, *Journal of Molecular Liquids*, 2008, **143**, 52–63.

47 Y. Nannoolal, J. Rarey, D. Ramjugernath and W. Cordes, *Fluid Phase Equilibria*, 2004, **226**, 45–63.

48 S. Stein and R. Brown, *Journal of Chemical Information and Computer Science*, 1994, 581–587.

49 M. H. Barley, D. Topping, D. Lowe, S. Utembe and G. McFiggans, *Atmospheric Chemistry and Physics*, 2011, **11**, 13145–13159.

50 K. Joback and R. Reid, *Chemical Engineering Communications*, 1987, **57**, 233–243.

51 A. M. Booth, M. H. Barley, D. O. Topping, G. McFiggans, A. Garforth and C. J. Percival, *Atmospheric Chemistry and Physics*, 2010, **10**, 4879–4892.

52 M. Capouet and J.-F. Muller, *Atmospheric Chemistry and Physics*, 2006, **6**, 1455–1467.

53 J. F. Pankow and W. E. Asher, *Atmospheric Chemistry and Physics*, 2008, **8**, 2773–2796.

54 J. L. Fry and K. Sackinger, *Atmospheric Chemistry and Physics*, 2012, **12**, 8797–8811.

55 J. Pankow, *Atmospheric Environment*, 1994, **28**, 185–188.

56 R. Valorso, B. Aumont, M. Camredon, T. Raventos-Duran, C. Mouchel-Vallon, N. L. Ng, J. H. Seinfeld, J. Lee-Taylor and S. Madronich, *Atmospheric Chemistry and Physics*, 2011, **11**, 6895–6910.

57 J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics*, John Wiley & Sons Inc., Hoboken, NJ, USA, 2nd edn, 2006.

58 C. D. Cappa and J. L. Jimenez, *Atmospheric Chemistry and Physics*, 2010, **10**, 5409–5424.

59 R. S. of Chemistry, *ChemSpider: Search and share chemistry*, 2014, `http://www.chemspider.com`.

60 N. I. for Science and Technology, *NIST Chemistry WebBook: NIST*

Standard Reference Database Number 69, 2014, `http://webbook.nist.gov/chemistry/name-ser.html`.

61 N. M. Donahue, A. L. Robinson and S. N. Pandis, *Atmospheric Environment*, 2009, **43**, 94–106.

62 K. Salo, A. M. Jonsson, P. U. Andersson and M. Hallquist, *The Journal of Physical Chemistry A*, 2010, **114**, 4586–94.

63 C. D. Cappa, E. R. Lovejoy and a. R. Ravishankara, *The Journal of Physical Chemistry A*, 2007, **111**, 3099–109.

64 W. E. Asher, J. F. Pankow, G. B. Erdakos and J. H. Seinfeld, *Atmospheric Environment*, 2002, **36**, 1483–1498.