Uncertain Henry's Law Constants Compromise Equilibrium Partitioning Calculations of Atmospheric Oxidation Products

Chen Wang[1], Tiange Yuan[1], Stephen A. Wood[1], Kai-Uwe Goss[2,3], Jingyi Li,[4,5] Qi Ying[4], Frank Wania[1*]

[1]*Department of Physical and Environmental Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON, M1C 1A4, Canada*

[2]*Department of Analytical Environmental Chemistry, Centre for Environmental Research UFZ Leipzig-Halle, Permoserstraße 15, Leipzig, D-04318, Germany*

[3]*Institute of Chemistry, University of Halle-Wittenberg, Kurt-Mothes-Straße 2, Halle, D-06120, Germany*

[4]*Environmental and Water Resources Division, Zachry Department of Civil Engineering, Texas A&M University College Station, TX 77843-3136*

[5]*Current Address: School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, 219 Ningliu Road, Nanjing 210044, China*

[*]*Corresponding author: frank.wania@utoronto.ca, +1-416-287-7225*

Abstract

Gas-particle partitioning governs the distribution, removal and transport of organic compounds in the atmosphere and the formation of secondary organic aerosol. The large variety of atmospheric species and their wide range of properties make predicting this partitioning equilibrium challenging. Here we expand on earlier work and predict gas-organic and gas-aqueous phase partitioning coefficients for 3414 atmospherically relevant molecules using COSMOtherm, SPARC and poly-parameter linear free energy relationships. The Master Chemical Mechanism generated the structures by oxidizing primary emitted volatile organic compounds. Predictions for gas-organic phase partitioning coefficients ($K_{WIOM/G}$) by different methods are on average within one order of magnitude of each other, irrespective of the numbers of functional groups, except for predictions by COSMOtherm and SPARC for compounds with more than three functional groups, which have a slightly higher discrepancy. Discrepancies between predictions of gas-aqueous partitioning ($K_{W/G}$) are much larger and increase with the number of functional groups in the molecule. In particular, COSMOtherm often predicts much lower $K_{W/G}$ for highly functionalized compounds than the other methods. While the quantum-chemistry based COSMOtherm accounts for the influence of intramolecular interactions on conformation, highly functionalized molecules likely fall outside of the applicability domain of the other techniques, which at least in part rely on empirical data for calibration. Further analysis suggests that atmospheric phase distribution calculations are sensitive to the partitioning coefficient estimation method, in particular to the estimated value of $K_{W/G}$. The large uncertainty in $K_{W/G}$ predictions for highly functionalized organic compounds needs to be resolved to improve the quantitative treatment of SOA formation.

Introduction

Volatile organic compounds (VOCs) emitted to the atmosphere are oxidized to form secondary products. These products tend to be more oxygenated, less volatile and more water-soluble than their parent compounds, and thus have higher affinity for aerosol particles and aqueous droplets. Equilibrium partitioning coefficients are often needed to assess the distribution of these oxidized compounds among different phases in the atmosphere such as aerosol particles, fog and cloud droplets. In particular, the partitioning between gas and organic phase and between gas and aqueous phase is required for the evaluation of an organic compound's contribution to secondary organic aerosol (SOA) formation, its transport, removal and lifetime. Experimentally determined partitioning coefficients are rarely available for the oxidation products of VOCs due to the difficulties in making the measurements and obtaining chemical standards. Furthermore, there are many thousand organic species in the atmosphere (Hallquist et al., 2009); the number is even higher when considering their isomers. Their gas-particle partitioning is therefore usually predicted. Reliable estimation methods for gas-organic and gas-aqueous partitioning should be applicable to a wide range of organic compounds, especially to multifunctional species generated during the multi-step atmospheric oxidation of precursor VOCs.

Current approaches for predicting partitioning into non-aqueous organic aerosol phases almost exclusively rely on predictions of vapor pressure. These predictions have large uncertainties; comparison among different vapor pressure prediction methods suggest increasing discrepancies with increasing numbers of functional groups in an organic compound (Valorso et al., 2011;Barley and McFiggans, 2010;McFiggans et al., 2010;Compernolle et al., 2011). This uncertainty matters, because it is the multi-functional oxidation products that can occur in either gas or condensed phases in the atmosphere. Instead of relying on predictions for vapor pressures, Wania et al. (2014) proposed using three alternative methods for direct gas-particle partitioning prediction: poly-parameter linear free energy relationships (ppLFERs), the on-line calculator of SPARC Performs Automated Reasoning in Chemistry (SPARC) and the quantum-chemistry based program COSMOtherm. Wania et al. (2014) found that partitioning coefficients

66 predicted for the oxidation products of n-alkanes are within one order of magnitude, and

67 mutual agreement does not deteriorate with increasing number of functional groups. Because

68 of the relatively small number of oxidation products in that study, the reliability of these

69 prediction methods for other organic compounds requires further evaluation.

70 While more experimental data exist for the Henry's law constant of atmospherically relevant

71 compounds than gas-organic phase partitioning coefficients (Sander, 2015), data are not usually

72 available for VOC oxidation products, which potentially have a higher affinity for atmospheric

73 aqueous phases and a great atmospheric abundance. Currently available prediction methods

74 for the air-water partitioning coefficient include GROup contribution Method for Henry's law

75 Estimate (GROMHE) (Raventos-Duran et al., 2010), SPARC (Hilal et al., 2008), HENRYWIN in EPI

76 suite (US EPA, 2012), and ppLFERs (Goss, 2006). Sander (2015) provides a more comprehensive

77 list of websites as well as quantitative structure-property relationships for Henry's law

78 constants. COSMOtherm can also predict gas-aqueous phase partitioning of organic compounds,

79 including VOC oxidation products (Wania et al., 2015). Though many different methods are

80 available for Henry's law constant prediction, they have not been systematically evaluated for a

81 large set of organic compounds of atmospheric relevance. An exception is the comparison of

82 GROMHE, SPARC and HENRYWIN predictions for 488 organic compounds bearing functional

83 groups of atmospheric relevance (Raventos-Duran et al., 2010).

84 Even for the relatively simple molecules for which experimental evaluation data exist, these

85 methods have considerable uncertainties. Raventos-Duran et al. (2010) reported Root Mean

86 Square Errors (RMSE) of 0.38, 0.61, and 0.73 log units for Henry's constants predicted by

87 GROMHE, SPARC and HENRYWIN, respectively. The ppLFER developed by Goss (2006) has a

88 RMSE of 0.15 log units for the 217 compounds used for calibration. The error can be expected

89 to be much larger for molecules that either are not part of the calibration (GROHME, ppLFER) or

90 are more complex. For a compound with multiple functional groups, Isaacman-VanWertz et al.

91 (2016) found discrepancies in predicted Henry's law constant of several orders of magnitude.

92 Hodzic et al. (2014)'s method of estimating the Henry's law constant for atmospheric oxidation

93 products of different precursors also has uncertainties of several orders of magnitude.

94 The objective of this paper was to compare and evaluate gas-particle partitioning predictions

95 for a large number of organic compounds of atmospheric interest using ppLFER (in combination

96 with ABSOLV-predicted solute descriptors), SPARC and COSMOtherm. While all three methods

97 are able to estimate both gas-organic and gas-aqueous partitioning, they are based on different

98 principles: ppLFERs are empirically calibrated multiple linear regressions, SPARC contains

99 solvation models based on fundamental chemical structure theory (Hilal et al., 2004), and

100 COSMOtherm combines quantum chemistry with statistical thermodynamics (Klamt and Eckert,

101 2000). This study thus expands earlier work (Wania et al., 2014) to a much larger number of

102 compounds and to aqueous phase partitioning. As such, it includes quantum-chemistry based

103 predictions for an unprecedented number of atmospherically relevant compounds.

Method

105 The Master Chemical Mechanism (MCM v3.2, *http://mcm.leeds.ac.uk/MCM*) a near-explicit

106 chemical mechanism was used to generate 3414 non-radical species through the multi-step gas

107 phase oxidation of 143 parent VOCs (methane + 142 non-methane VOCs). Reactions of the

108 parent VOCs with O_3, OH and NO_3 are included in the MCM mechanism whenever such

109 reactions are possible. The details about the studied compounds are given in the supporting

110 information (Excel spreadsheet), including the compounds' MCM ID, SMILES, precursors (i.e.

111 the parent VOC), molecular weight, molecular formula, elements, generation of oxidation,

112 number and species of functional groups, O:C ratio, and average carbon oxidation state ($\overline{OS}_C$)

113 (Kroll et al., 2011).

114 Three prediction methods are used to estimate the equilibrium partitioning coefficients

115 between a water-insoluble organic matter phase (WIOM) and the gas phase ($K_{WIOM/G}$) at 15 °C in

116 units of m^3 (air)/m^3 (WIOM) as well as the equilibrium partitioning coefficients between water

117 and gas phase ($K_{W/G}$) at 15 °C in units of m^3 (air)/m^3 (water). The two partitioning coefficients

118 are defined as:

119 $K_{WIOM/G} = C_{WIOM}/C_G$ (1)

120 $K_{W/G} = C_W/C_G$ (2)

121 C_{WIOM}, C_W and C_G (mol/m³) are equilibrium concentrations of an organic compound in WIOM,

122 water, and gas phase, respectively. Partitioning between gas and aqueous phase can be

123 significantly influenced by the presence of inorganic salts (i.e. the salt effect) (Endo et al.,

124 2012;Wang et al., 2016;Wang et al., 2014;Waxman et al., 2015), the hydration of carbonyls (Ip

125 et al., 2009) and the dissociation of organic acids (Mouchel-Vallon et al., 2013), particularly in

126 the aqueous phase of aerosols. However, in this study only the partitioning between gas and

127 pure water, i.e. the Henry's law constant, is predicted, and no hydration, salt effect or acid

128 dissociation is considered. Conversion of partitioning coefficients $K_{W/G}$ to Henry's constant (K_H)

129 in units M/atm or $K_{WIOM/G}$ to saturation concentration (C^*, µg/m³) is provided in the supporting

130 information.

131 Wania et al. (2014) describe each prediction method in detail. In brief, ppLFERs are developed

132 by performing a multi-linear regression of experimental K values against compound specific

133 solute descriptors (Endo and Goss, 2014). These descriptors represent a solute's hydrogen-

134 bond acidity (A), hydrogen-bond basicity (B), dipolarity/polarizability (S), McGowan volume

135 (cm³/mol) divided by 100 (V), excess molar refraction (E), and logarithmic hexadecane-air

136 partitioning constant at 25°C (L). In this study, solute descriptors for the 3414 compounds were

137 predicted with ABSOLV (ACD/Labs, Advanced Chemistry Development, Inc., Toronto, Canada).

138 The regression coefficients in ppLFERs are denoted by a, b, s, v, e, and l; c is the regression

139 constant. The ppLFER for air-water partitioning was taken from (Goss, 2006):

140 $\log K_{W/G} = c + aA + bB + sS + vV + eE$ (3)

141 whereas ppLFERs for four different organic aerosol were taken from (Arp et al., 2008);

142 $\log K_{Aerosol/G} = c + aA + bB + sS + vV + lL$ (4)

143 As described in Wania et al. (2014), the average of the four $K_{Aerosol/G}$ was compared with the

144 $K_{WIOM/G}$ predicted by the other two methods. SPARC is a commercial web-based calculator for

145 prediction of physical chemical properties from molecular structure developed by the US

146 Environmental Protection Agency (Hilal et al., 2004). The predictions of $K_{W/G}$ and $K_{WIOM/G}$ are

147 based on solvation models in SPARC that describe the intermolecular interaction between

different molecules (solute and solvent), including dispersion, induction, dipole-dipole, and H-bonding interactions, which are developed and calibrated with experimental data (Hilal et al., 2008). For the calculations of $K_{WIOM/G}$ by SPARC and COSMOtherm, the phase WIOM is represented by the surrogate structure "B" as proposed by Kalberer et al. (2004) and adopted previously by Arp and Goss (2009) and Wania et al. (2014). SPARC calculations were carried out using the on-line calculator (http://archemcalc.com/sparc-web/calc), with SMILES (simplified molecular-input line-entry system) strings as input. COSMOtherm predicts a large variety of properties based on COSMO-RS (conductor-like screening model for real solvents) theory, which uses quantum-chemical calculations and statistical thermodynamics (Klamt and Eckert, 2000;Klamt, 2005). First, TURBOMOLE (version 6.6, 2014, University of Karlsruhe & Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007 available from www.turbomole.com) optimizes the geometry of the molecules of interest at the BP-TZVP level. COSMOconf (version 3.0, COSMOlogic) then selects a maximum of ten lowest energy conformers for each calculated molecule and generates COSMO files. Calculations with TURBOMOLE and COSMOconf were performed on the General Purpose Cluster (GPC) supercomputer at the SciNet HPC Consortium at University of Toronto (Loken et al., 2010). Finally, COSMOtherm (version C30_1501 with BP_TZVP_C30_1501 parameterization, COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2015) calculates partitioning coefficients from the selected COSMO files at 15 °C.

In order to compare different predictions numerically, we calculated the mean difference (MD) and the mean absolute difference (MAD) for each pair of $K_{WIOM/G}$ or $K_{W/G}$ sets:

$$\mathrm{MD}_{XY} = \frac{1}{n}\sum_i \left(\log_{10}K_{i,CP/G\,X} - \log_{10}K_{i,CP/G\,Y}\right) \qquad (5)$$

$$\mathrm{MAD}_{XY} = \frac{1}{n}\sum_i \left|\log_{10}K_{i,CP/G\,X} - \log_{10}K_{i,CP/G\,Y}\right| \qquad (6)$$

where CP ("condensed phase") stands for either WIOM or water and X and Y represents two prediction techniques.

Results

The Range of Estimated Partitioning Coefficients

Partitioning coefficients predicted for each compound with different methods are given in an Excel spreadsheet as Supporting Information. All three methods predicted the log $K_{WIOM/G}$ for these organic compounds to range from approximately 0 to 15 (Figure 1 (a)-(c)). Hodzic et al. (2014) predicted a log $K_{WIOM/G}$ in the range of approximately 0 and 20 at 25 °C (see conversion between C^* and $K_{WIOM/G}$ in the supporting information) for oxidation products of different VOCs (including n-alkanes, benzene, toluene, xylene, isoprene and terpenes), i.e. their data set included higher $K_{WIOM/G}$ values (indicating generally lower volatility) than those generated here, even though $K_{WIOM/G}$ values are lower at higher temperature.

The log $K_{W/G}$ range predicted for the studied compounds by the three methods is more variable (Figure 1 (d)-(f)), with the ABSOLV/ppLFER predictions covering a wider range (-1.4 to 21.3) than either SPARC (-2.7 to 17.2) or COSMOtherm (-2 to 13.8). Hodzic et al. (2014) predicted a log $K_{W/G}$ in the range of -2.6 and 17.4 at 25°C (see conversion to K_H in the supporting information). The wider range of the ABSOLV/ppLFER predictions is due to much higher predicted $K_{W/G}$-values for compounds with the highest affinity for the aqueous phase.

Comparison between Different Prediction Methods

The discrepancies between different predictions (MAD and MD) are given in Table 1. Figure S1 in the supporting information illustrates the frequency of the discrepancies between different pairs of predicted log $K_{WIMO/G}$ and log $K_{W/G}$ values. This discrepancy only indicates the agreement between any two predictions with little indication of the accuracy of the prediction, for reasons discussed later. The agreement between the $K_{WIOM/G}$ predictions by COSMOtherm, SPARC and ABSOLV/ppLFER was reasonable (Figure 1 (a)-(c)). In particular, the MAD between $K_{WIOM/G}$ predictions is less than 1 log units (Table 1) and therefore similar to what had been previously found for a much smaller set of n-alkane oxidation products (Wania et al., 2014). The $K_{WIOM/G}$-values predicted by SPARC tend to be higher than those predicted by COSMOtherm and ABSOLV/ppLFER (MD of -0.64 and -0.79 in log units, respectively), whereas the latter two

predictions have a slightly better agreement, with a MD of 0.15 log units (Figure 1 (c) and Table 1). Overall, the agreement in the $K_{WIOM/G}$ predicted with these three methods, which are based on very different theoretical foundations, is much better than that between different vapor pressure estimation methods commonly used for gas-particle partitioning calculations (Valorso et al., 2011).

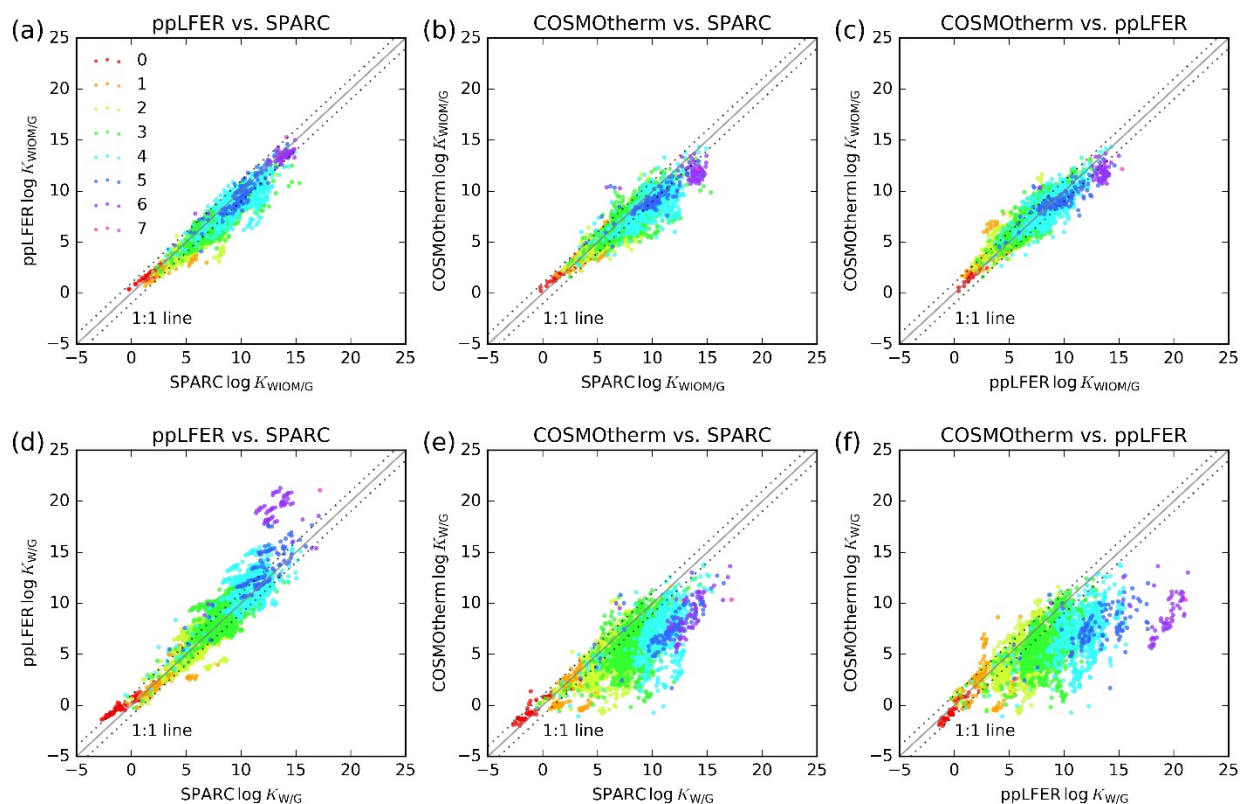



Figure 1   Comparison of the $K_{WIOM/G}$ (upper panel) and $K_{W/G}$ (lower panel) predicted using COSMOtherm, SPARC and ABSOLV/ppLFERs. The differently colored dots indicate the number of functional groups in the molecules. The solid line indicates a 1:1 agreement. The dotted lines indicate a deviation by ±1 log unit.

The $K_{W/G}$ predicted by ABSOLV/ppLFER and SPARC differ from COSMOtherm predictions substantially, on average by more than two orders of magnitude. In Figure 1 (e) and (f), predictions are more scattered (indicating a larger MAD) and most markers are located above the 1:1 line, indicating that $K_{W/G}$ predicted by COSMOtherm are mostly lower than those predicted by SPARC and ABSOLV/ppLFER, with a MD of -2.06 and -2.42 log units, respectively.

215 These discrepancies tend to increase with the $K_{W/G}$. Raventos-Duran et al. (2010) also showed

216 that the reliability of $K_{W/G}$ estimates made by GROMHE, SPARC and HENRYWIN decreases with

217 increasing affinity for the aqueous phase. $K_{W/G}$ predictions by SPARC and ABSOLV/ppLFER are

218 more consistent (with a MAD around 1 log units, see Figure 1 (d)). The largest discrepancies

219 between ABSOLV/ppLFER and SPARC (and also between ABSOLV/ppLFER and COSMOtherm)

220 occur for compounds with the highest $K_{W/G}$ as predicted by ABSOLV/ppLFER (purple markers in

221 Figure 1 (d) and (f)). Further analysis indicates that these compounds have the largest number

222 of functional groups (≥6) and oxygen (9~12 oxygen) in the molecule; this will be discussed in

223 detail below.

224 **Table 1** Mean absolute differences (MAD) and mean differences (MD) between SPARC,

225 ABSOLV/ppLFER and COSMOtherm predictions for compounds with different numbers

226 of functional groups

Number of Functional Groups			0	1	2	3	4	5	>5	All
Number of Compounds			63	372	1179	1064	565	111	60	3414
$\log K_{WIOM/G}$	ppLFER vs. SPARC	MAD	0.24	0.70	0.95	0.93	1.08	0.75	0.54	0.91
		MD	0.05	-0.70	-0.91	-0.81	-0.83	-0.24	-0.30	-0.79
	COSMOtherm vs. SPARC	MAD	0.36	0.48	0.80	0.94	1.42	1.22	2.11	0.94
		MD	0.29	-0.19	-0.55	-0.57	-1.21	-0.78	-1.83	-0.64
	COSMOtherm vs. ppLFER	MAD	0.32	0.67	0.63	0.74	0.89	0.93	1.72	0.73
		MD	0.24	0.51	0.36	0.24	-0.38	-0.54	-1.53	0.15
$\log K_{W/G}$	ppLFER vs. SPARC	MAD	0.75	0.57	0.84	1.08	1.48	1.53	5.78	1.10
		MD	0.74	-0.09	-0.15	0.38	0.87	1.45	5.76	0.36
	COSMOtherm vs. SPARC	MAD	0.51	0.86	1.61	2.31	3.78	4.34	4.55	2.23
		MD	0.48	-0.59	-1.44	-2.18	-3.74	-4.04	-4.36	-2.06
	COSMOtherm vs. ppLFER	MAD	0.40	1.16	1.64	2.63	4.62	5.55	10.09	2.64
		MD	-0.26	-0.50	-1.29	-2.56	-4.61	-5.50	-10.05	-2.42

227 **Dependence of Partitioning Coefficients on Attributes of the Compounds**

228 The equilibrium partitioning coefficients depend on molecular attributes. Here we explored this

229 dependency on the number of functional groups, molecular mass, generation of oxidation,

230 number of oxygens and O:C ratio.

231 Previous work observed that discrepancies between vapor pressure predictions by different

232 methods increased with the number of functional groups in atmospherically relevant organic

233 compounds (Valorso et al., 2011;Barley and McFiggans, 2010;Compernolle et al., 2011). For

234 instance, the MAD between different vapor pressure predictions increased from 0.47 to 3.6 log

235 units when the number of functional groups in the molecules increased from one to more than

236 three (Valorso et al., 2011). In order to explore if the partitioning coefficients predicted with

237 SPARC, ABSOLV/ppLFER and COSMOtherm show the same dependence on the number of

238 functional groups, we counted the number of hydroxyl (ROH), aldehyde (RCHO), ketone (RCOR'),

239 carboxylic acid (RCOOH), ester (RCOOR'), ether (ROR'), peracid (RCOOOH), peroxide (ROOH,

240 ROOR'), nitrate (NO3), peroxyacyl nitrate (PAN), nitro (NO2) groups, halogen (Cl, Br), and

241 sulphur (S) in the 3414 molecules. About two thirds (2243) of the compounds contain two or

242 three functional groups (Table 1). 736 compounds contain more than three functional groups

243 and the rest contains just one or no functional group. In Figure 1 the compounds are colored

244 according to the number of functional groups in a molecule and Table 1 lists the MAD and MD

245 between predictions based on the number of functional groups. The predicted partitioning

246 coefficients (both $K_{\text{WIOM/G}}$ and $K_{\text{W/G}}$) generally increase with the number of functional groups

247 (Figure 1 and Figure S2). Compounds with no functional groups are the precursor compounds,

248 which generally have a smaller discrepancy among different prediction methods.

249 The boxplots in Figure 2 show the difference in SPARC, ABSOLV/ppLFER and COSMOtherm

250 predictions for compounds having different number of functional groups. The mean absolute

251 difference in predicted log $K_{\text{WIOM/G}}$ is mostly (and on average) smaller than one log unit for

252 compounds with up to seven functional groups (Table 1). There is a slightly larger discrepancy in

253 the predicted log $K_{\text{WIOM/G}}$ values for compounds with more than three functional groups. The

254 agreement among different methods does not deteriorate as much with increasing number of

255 functional groups as that among vapor pressure predictions. The largest MADs of 1.72 and 2.11

256 between COSMOtherm and ABSOLV/ppLFER, and between COSMOtherm and SPARC,

257 respectively, for compounds with >5 functional groups (Table 1) are still much lower than

258 discrepancies reported between different vapor pressure prediction methods (Valorso et al.,

259 2011).

260   Different from the predictions for $K_{WIOM/G}$, the discrepancy between COSMOtherm and SPARC

261   and between COSMOtherm and ABSOLV/ppLFER in the predicted $K_{W/G}$ increases significantly

262   with the number of functional groups (Figures 1 and 2), from less than one order of magnitude

263   for compounds with no functional groups to up to five orders of magnitude for compounds with

264   more than three functional groups (Table 1). In addition, the MD in Table 1 and Figure 2

265   indicate that the discrepancies are almost always in one specific direction, i.e. a lower value of

266    $K_{W/G}$ estimated by COSMOtherm. This is evidenced by the almost identical absolute values of

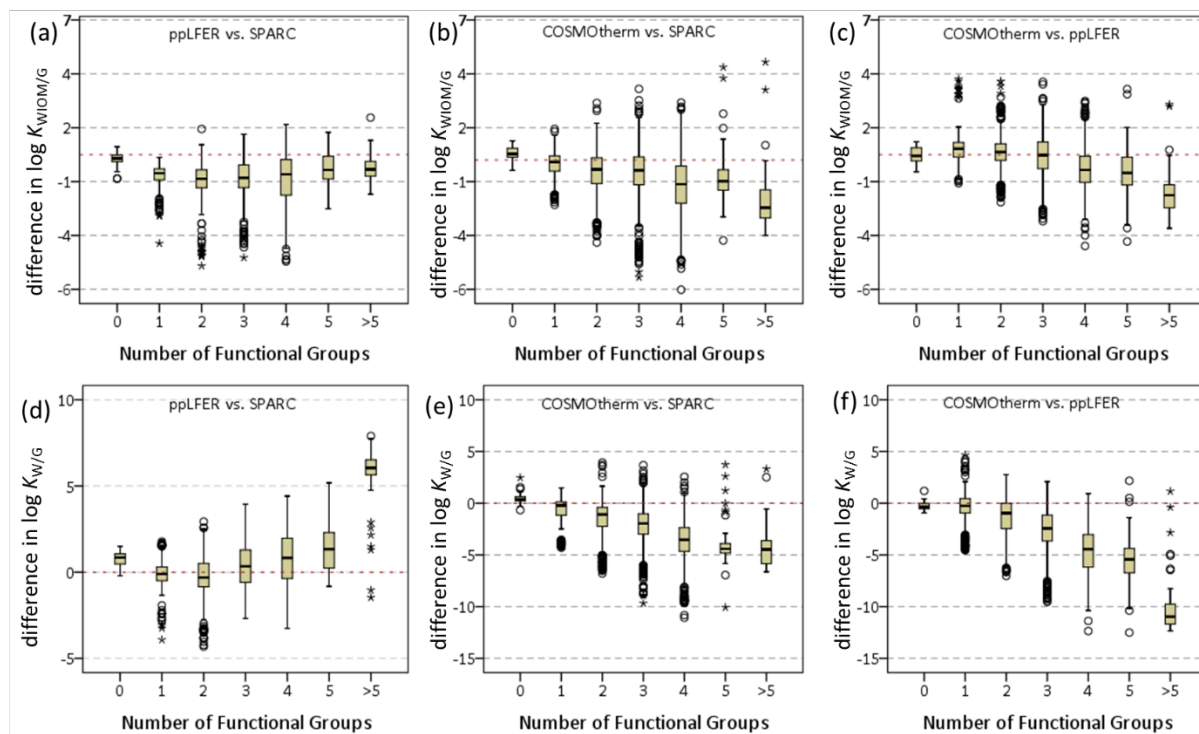
267   MAD and MD between COSMOtherm and ABSOLV/ppLFER and between COSMOtherm and

268   SPARC for compounds with more than three functional groups (Table 1). The uncertainty of the

269   SPARC, ABSOLV/ppLFER and COSMOtherm predictions of $K_{W/G}$ tends to increase with the

270   number of functional groups. Clearly, the reliability of $K_{W/G}$ estimates for multifunctional

271   compounds needs further assessment.

272

273   **Figure 2**   Boxplot of difference in SPARC, ABSOLV/ppLFER and COSMOtherm predictions for

274   compounds with different number of functional groups. The line inside each box shows

275   the median difference for log $K_{WIOM/G}$ or log $K_{W/G}$ for different categories of compounds.

276   The marker circle and star indicates possible outliers and extreme values, respectively.

277   Note the different scales for different panels.

278   It is also possible to explore the dependence of the prediction discrepancy on other molecular

279   attributes, such as molecular mass (Figures S3 and S4), the number of oxygen in the molecule

280   (Figures S5 and S6), the O:C ratio (Figure S7), the number of oxidation steps a molecular has

281   undergone (oxidation generation, Figure S8), or the number of occurrences of a specific type of

282   functional group, e.g. hydroxyl, in a molecule (Figure S9). The prediction discrepancies become

283   larger with an increase in each of these parameters, especially for $K_{W/G}$. This is not surprising as

284   these molecular attributes all tend to be highly correlated, i.e. with each oxidation step a

285   molecule becomes more oxygenated, has a large molar mass, a larger number of oxygen, a

286   higher O:C ratio, and a larger number of functional groups.

Discussion

288   We believe there are primarily two factors that are contributing to errors in the prediction of

289    $K_{CP/G}$ for the SOA compounds. One is the lack of experimental data for compounds that are

290   similar to the SOA compounds, which implies that prediction methods relying on calibration

291   with experimental data are being used outside their applicability domain. The other is the

292   failure of some prediction methods to account for the various conformations that compounds

293   with multiple functional groups can undergo due to extensive intra-molecular interaction

294   (mostly internal hydrogen bonding, see Figure S10 for example). The two factors are related: in

295   some instances a prediction method cannot account for such conformations precisely because

296   the calibration data set does not contain compounds that undergo such intra-molecular

297   interactions.

298   SPARC relies to some extent on calibrations with empirical data. While the experimental data

299   underlying SPARC have not been disclosed, it is highly unlikely that they include multifunctional

300   compounds of atmospheric relevance (e.g. compounds containing multiple functional groups,

301   including peroxides, peroxy acids etc.), simply because such empirical data do not exist. It is

302   therefore safe to assume that many of the 3414 SOA compounds will fall outside of the domain

303 of applicability of SPARC. It is also likely that SPARC can only account for intra-molecular

304 interactions and conformations to a limited extent, if at all.

305 In the case of ppLFER, there are actually two predictions that rely on calibration with empirical

306 data, the prediction of solute descriptors and the prediction of $K_{CP/G}$. The solute descriptors are

307 predicted with ABSOLV, because experimentally measured descriptors are unavailable for

308 multifunctional atmospheric oxidation products. ABSOLV relies on a group contribution

309 approach (Platts et al., 1999) complemented by some other, undisclosed procedures that make

310 use of experimental partitioning coefficients between various phases (ACD/Labs, 2016). Again,

311 those experimental data do not comprise compounds structurally similar to the multifunctional

312 atmospheric oxidation products considered here. As a group contribution method, which adds

313 up the contributions of different functional groups to a compound's property, ABSOLV

314 therefore cannot, or only to a limited extent, consider the interactions between different

315 functional groups in a molecule.

316 Ideally, when supplied with well-characterized solute descriptors, ppLFERs should be able to

317 consider the influence of both intra-molecular interactions and the interactions a molecule has

318 with its surroundings, i.e. the involved partitioning phases. Even if a molecule has different

319 conformations in different phases, i.e. if the solute descriptors for a compound are phase

320 dependent, it is possible to derive well-calibrated "average" descriptors to use in a ppLFER

321 (Niederer and Goss, 2008). However ABSOLV cannot correctly predict such "average"

322 descriptors and our ppLFER predictions therefore cannot account for the influence of

323 conformations.

324 In the case of the actual ppLFER prediction of $K_{W/G}$ and $K_{WIOM/G}$, the empirical calibration

325 datasets are public (Goss, 2006;Arp et al., 2008) and do not comprise compounds that are

326 representative of the 3414 SOA compounds in terms of the number of functional groups per

327 molecule or the range of K-values. For instance, the log $K_{W/G}$ of the 217 compounds Goss (2006)

328 used for the development of a ppLFER ranged from -2.4 to 7.4, i.e. the highest $K_{W/G}$ predicted

329 here is almost 14 orders of magnitude higher than the highest $K_{W/G}$ included in the calibration.

330 Similarly, Arp and Goss (2009) developed the ppLFERs for atmospheric aerosol from an

331   empirical dataset of 50~59 chemicals, whose log $K_{WIOM/G}$ ranged from approximately 2 to 7. The

332   highest $K_{WIOM/G}$ predicted here is eight orders of magnitude higher. Predictions for compounds

333   outside of the calibration domain may introduce large errors and the high $K_{W/G}$ and $K_{WIOM/G}$

334   values estimated by ppLFER can thus be expected to be highly uncertain. Overall, however, we

335   expect the uncertainty of the ABSOLV-predicted solute descriptors to be larger than the

336   uncertainty introduced by the ppLFER equation, especially for the relatively well-calibrated

337   water/gas phase partition system. While the use of measured solute descriptors therefore

338   would likely greatly improve the ppLFER prediction (Endo and Goss, 2014), those are unlikely to

339   become available for atmospheric oxidation products.

340   In contrast to the other methods, COSMOtherm relies only in a very fundamental way on some

341   empirical calibrations (and these calibrations are not specific for specific compound classes or

342   partition systems) and it considers intra-molecular interactions and the different conformations

343   of a molecule. As such, COSMOtherm is not constrained by the limitations the other methods

344   face, namely the lack of suitable calibration data, which necessitates extreme extrapolations

345   and predictions beyond the applicability domain, and the failure to account for the effect of

346   intra-molecular interactions and conformations on the interactions with condensed phases.

347   Because intra-molecular interactions are likely to reduce the potential of a compound to

348   interact with condensed phases (i.e. the organic and aqueous phase), ignoring them can be

349   expected to lead to overestimated partitioning coefficients $K_{CP/G}$ and to underestimated vapor

350   pressures (P_L) and C^*, i.e. underestimating the volatility of the organic compounds. This is

351   consistent with COSMOtherm-predicted $K_{WIOM/G}$ and $K_{W/G}$-values for multifunctional compounds

352   that are lower than the SPARC and ABSOLV/ppLFER predictions (i.e. MD<0 in Table 1), because

353   the latter do not account for the influence of intra-molecular interactions. Kurtén et al. (2016)

354   similarly found that COSMOtherm-predicted saturation vapor pressures for most of the more

355   highly oxidized monomers were significantly higher (up to 8 orders of magnitude) than those

356   predicted by group-contribution methods. The wider range on the higher end of the log C^*

357   values estimated by Hodzic et al. (2014) is possibly due to the large uncertainties associated

358   with vapor pressure estimation (likely underestimation) for low volatile compounds. Valorso et

359 al. (2011) also found group contribution methods to underestimate the saturation vapor
360 pressure of multifunctional species.

361 Compared to $K_{WIOM/G}$, P_L and C^*, ignoring intra-molecular interaction is likely even more
362 problematic in the case of $K_{W/G}$ prediction. Intra-molecular interactions mostly affect the ability
363 of the molecule to undergo H-bonding with solvent molecules. The system constants describing
364 H-bond interactions (a and b) are larger in the ppLFER equations for $K_{W/G}$ than in the one for
365 $K_{WIOM/G}$ (Arp et al., 2008;Goss, 2006), indicating a stronger effect of H-bonds on water/gas
366 partitioning than WIOM/gas partitioning. This likely is the reason why the COSMOtherm-
367 predicted $K_{W/G}$ are so much lower than the $K_{W/G}$ predicted by the other two methods, whereas
368 the difference is much smaller for the $K_{WIOM/G}$ (Table 1). It likely also explains why the
369 discrepancies among the predicted $K_{W/G}$ increase with the number of functional groups. It is
370 more difficult to predict $K_{W/G}$ than $K_{WIOM/G}$, because the free energy cost of cavity formation in
371 water is influenced more strongly by H-bonding and therefore much more variable than in
372 WIOM. Certainly, the activity coefficient in water (γ_W) is much more variable than the activity
373 coefficient in WIOM (γ_{WIOM}) for the investigated substances. log γ_{WIOM} predicted by
374 COSMOtherm at 15 °C varies from -3.8 to 1.8 (with an average of 0.04, indicating a γ_{WIOM} close
375 to unity, and a standard deviation of 0.5, 94 % of the compounds have a log γ_{WIOM} between -1
376 and 1), whereas γ_W ranges from -2.3 to 8.9 (with an average of 2.7 and a standard deviation of
377 1.4) (Supporting information Excel spreadsheet and Figure S11).

378 In the absence of experimental data for multi-functional SOA compounds, we do not know
379 whether COSMOtherm-predicted $K_{W/G}$ and $K_{WIOM/G}$ values are any better than the other
380 predictions. For example, two earlier studies suggested that COSMOtherm might be
381 overestimating vapor pressures of multi-functional oxygen-containing compounds (Kurtén et al.,
382 2016;Schröder et al., 2016). However, we can infer that:

383 - the fact that COSMOtherm on the one hand and ABSOLV/ppLFERs and SPARC on the other
384   hand predict $K_{WIOM/G}$ that are on average within one order of magnitude for all studied
385   compounds, and less than two orders of magnitude for highly oxygenated multifunctional
386   organic compounds, lends credibility to all three predictions and suggests that partly

387   ignoring intra-molecular interactions and extrapolating beyond the applicability domain

388   incurs only limited errors in the $K_{WIOM/G}$ prediction of ABSOLV/ppLFERs and SPARC. In

389   addition, COSMOtherm and SPARC use a single surrogate molecule to represent the WIOM

390   phase, while ppLFERs were calibrated from atmospheric aerosols. The agreement among

391   different methods suggests that the surrogate suitably represents the solvation properties

392   of organic aerosol.

393 - the generally better agreement between $K_{W/G}$ values predicted by ABSOLV/ppLFER and

394   SPARC (Figure 1(d)) should not be seen as an indication that these methods are better at

395   predicting $K_{W/G}$. In fact, the lower $K_{W/G}$ values predicted by COSMOtherm have a higher

396   chance of being correct than the $K_{W/G}$ values predicted by ABSOLV/ppLFER and SPARC.

397 While ABSOLV/ppLFERs, SPARC and the group contributions methods currently used in the

398 atmospheric chemistry community are much more easily implemented for the large number of

399 compounds implicated in SOA formation, the current study demonstrates that the expertise

400 and time required to perform quantum-chemical calculations for atmospherically relevant

401 molecules should constitute but a minor impediment to a wider adoption of COSMOtherm

402 predictions. Here, we are not only compiling all the predictions we have made in the supporting

403 information file, we are also making available the cosmo-files (see Data Availability for details),

404 whose generation is the major time and CPU-demanding step in the use of COSMOtherm.

Atmospheric Implications

406 The phase distribution of an organic compound in the atmosphere depends on its partitioning

407 coefficients. The two-dimensional partitioning space defined by log $K_{W/G}$ and log $K_{WIOM/G}$

408 introduced recently (Wania et al., 2015) is used here to illustrate the difference in the

409 equilibrium phase distribution of these compounds in the atmosphere that arises from using

410 partitioning coefficients estimated by different methods (Figure 3). A detailed description of

411 partitioning space has been provided by Wania et al. (2015), a brief explanation is given in the

412 supporting information (Figure S12). Briefly, the blue solid lines between the differently colored

413 fields indicate partitioning property combinations that lead to equal distributions between two

414 phases in a phase-separated aerosol scenario, with a liquid water content (LWC) of 10 µg/m^3

415   and organic matter loading (OM) of 10 µg/m^3. The blue dotted lines represent a cloud scenario

416   where LWC is 0.3 g/m^3 and OM is 10 µg/m^3. Figures S13 and S14 in the supporting information

417   show an aerosol scenario without an aqueous phase and a cloud scenario without a separated

418   organic phase because all of the OM is dissolved in the aqueous phase (see also Figure S12 (c)

419   and (d)). Compounds are located in the partitioning space based on their estimated partitioning

420   coefficients ($K_{WIOM/G}$ and $K_{W/G}$). Compounds on the boundary lines have 50 % in either of the

421   two phases on both sides of the boundary and are thus most sensitive to uncertain partitioning

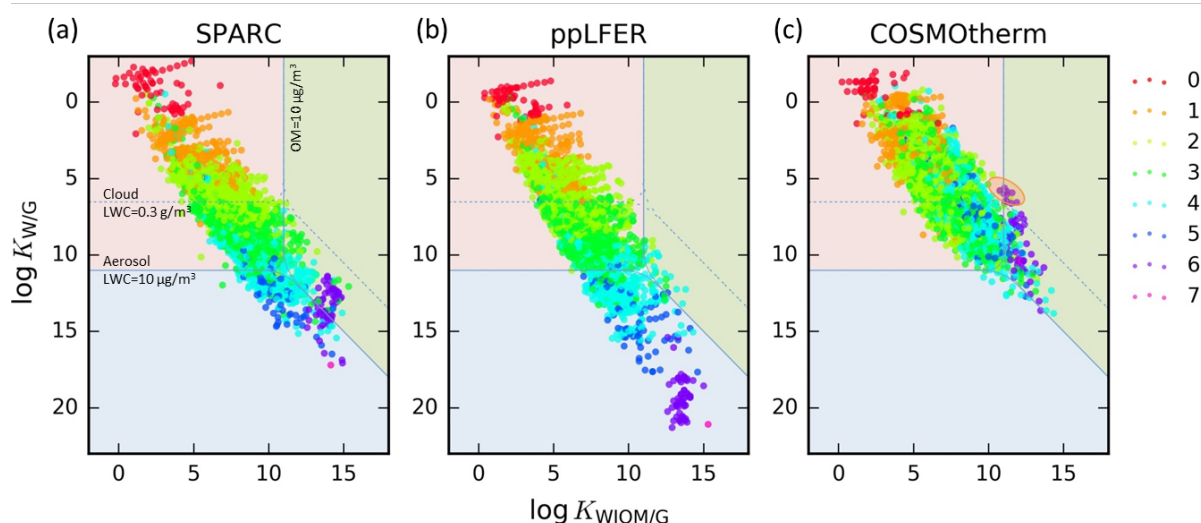
422   properties. On the other hand, for substances that fall far from the boundary lines indicating a

423   phase transition (e.g. volatile compounds with two or less functional groups), even relatively

424   large uncertainties in the partitioning coefficients could be tolerated, because they are

425   inconsequential.

426

427

428   **Figure 3**   Partitioning space plot, showing in pink, blue and green the combinations of partitioning

429   properties that lead to dominant equilibrium partitioning to the gas, aqueous, and

430   WIOM phases, respectively. The blue solid and dotted lines are boundaries for an

431   aerosol scenario (LWC 10 µg/m^3, 10 µg/m^3 OM) and a cloud scenario (LWC 0.3 g/m^3, 10

432   µg/m^3 OM), respectively. The differently colored dots indicate the number of functional

433   groups in the molecules.

434   When plotted in the chemical partition space, the 3414 chemicals occupy more or less the same

435   region as the much smaller set of SOA compounds investigated earlier (Wania et al., 2015).

436   When using predictions by COSMOtherm the SOA compounds cover a relatively smaller region

437   as compared to ABSOLV/ppLFER and SPARC. With increasing number of functional groups

438   (Figure 3) or molecular weight (Figure S15), an increasing fraction of these compounds

439   partitions into the condensed phases, i.e. WIOM or water. In general, compounds with water or

440   WIOM as the dominant phase usually are multifunctional, i.e. contain more than two functional

441   groups. According to Figure S15, compounds with predominant partitioning into WIOM usually

442   have a molar mass in excess of 200 g/mol, while some compounds with molar mass less than

443   200 g/mol prefer the aqueous phase. Other than the water content and WIOM loadings

444   illustrated in Figure 3, in reality a compound's atmospheric phase distribution depends on other

445   factors such as the organic matter composition, salt content, pH, and temperature (Wania et

446   al., 2015;Wang et al., 2015).

447   Comparing the different panels of Figure 3 reveals that the atmospheric equilibrium phase

448   distribution of SOA compounds can be very different depending on which methods is used for

449   partitioning coefficient estimation. The difference is most striking when comparing the

450   placement of highly functionalized compounds (with more than 3 functional groups) based on

451   ABSOLV/ppLFER and COSMOtherm predictions. The large $K_{W/G}$ values estimated by

452   ABSOLV/ppLFERs lead to these compounds having a high affinity for aqueous aerosol. In

453   contrast, predictions by COSMOtherm suggest that only very few of them (and not even the

454   ones with the highest number of functional groups) prefer the aqueous aerosol phase; instead

455   most of them have either gas or WIOM as the dominant phase. SPARC predicts a slightly larger

456   preference of highly functionalized compounds for the aqueous phase than COSMOtherm.

457   In a cloud scenario with a much higher LWC (shown by the blue dotted boundary lines in Figure

458   3), the choice of $K_{W/G}$ prediction method also matters. Whereas with ABSOLV/ppLFER and

459   SPARC most of the highly functionalized compounds (i.e. 96 % or 97 % of the 736 compounds

460   with >3 functional groups) partitions into aqueous phase, only two-thirds (64 %) do so when the

461    $K_{W/G}$s predicted by COSMOtherm are used. Further, only COSMOtherm predicts that some of

462   the SOA compounds (circled in Figure 3 (c)) would prefer to form a separate WIOM phase

463   rather than dissolve in the bulk aqueous phase. Those compounds are not sufficiently soluble in

464   water to partition to the cloud and are not sufficiently volatile to be in the gas phase.

465   Table 2 summarizes the number and percentage of compounds that have dominant partitioning

466   (at least 50 %) into different phases, which shows the impact of using different prediction

467   techniques on phase distribution calculations in different atmospheric scenarios. In a

468   parameterisation of SOA formation that includes an aqueous aerosol phase, use of $K_{W/G}$

469   predicted by ABSOLV/ppLFERs (and probably also the commonly employed group contribution

470   methods) would lead to much higher SOA mass than use of $K_{W/G}$ predicted by COSMOtherm.

471   For instance, 10 % and 17 % of the compounds predominantly partition into the aqueous phase

472   when predictions by SPARC and ABSOLV/ppLFER are used, in contrast to only 14 compounds

473   (less than 1 %) with COSMOtherm predictions (Table 2 scenario (a)). A large difference also

474   occurs in the cloud scenarios (Table 2 scenarios (b) and (d)), where SPARC and ABSOLV/ppLFER

475   predict twice as many compounds partitioning into the aqueous phase than COSMOtherm.

476   Incidentally, in a parameterization of SOA formation that does not account for an aqueous

477   aerosol phase (the scenario in Figure S12 (c) and Table 2 (c)), the impact of the choice of

478   partitioning prediction method is much smaller. The number of compounds on the right side of

479   the blue dotted boundary in Figure S13 does not vary substantially with different predictions.

480   Table S1 in the supporting information summarizes the number and percentage of compounds

481   that change their partitioning between gas and condensed phase under different atmospheric

482   conditions when a different prediction method is used. Depending on the scenarios, a total of

483   2.0 % up to 34 % of the 3414 compounds have a different dominant phase when using a

484   different prediction method. This change is larger for the cloud scenarios and much lower for

485   the aerosol scenarios especially if the aerosol contains no water.

486   **Table 2**   Percentage and number of compounds with at least 50 % in gas, water or WIOM phase

487   under different aerosol and cloud scenarios predicted with SPARC, ABSOLV/ppLFER and

488   COSMOtherm. The four scenarios (a-d) correspond to the scenarios in Figure S12 (a-d) in

489   Supporting information.

aerosol scenarios	(a) (LWC=10 µg/m³, OM=10 µg/m³)			(c) without water phase (LWC=0 µg/m³, OM=10 µg/m³)	
	Φ_G>50 %[a]	Φ_W>50 %[a]	Φ_{WIOM}>50 %[a]	Φ_G>50 %	Φ_{WIOM}>50 %
SPARC	85 % (2892)[b]	10 % (352)	4% (134)	92 % (3132)	8 % (282)
ABSOLV/ppLFER	82 % (2804)	17 % (570)	1% (25)	96 % (3267)	4 % (141)
COSMOtherm	96 % (3268)	0 % (14)	3% (119)	96 % (3282)	4 % (131)
cloud scenarios	(b) (LWC=0.3 g/m³, OM=10 µg/m³)			(d) without WIOM phase (LWC=0.3 g/m³, OM=0 µg/m³)	
	Φ_G>50%	Φ_W>50 %	Φ_{WIOM}>50 %	Φ_G>50 %	Φ_W>50 %
SPARC	36 % (1242)	64 % (2168)	0 % (0)	36 % (1242)	64 % (2172)
ABSOLV/ppLFER	35 % (1201)	65 % (2211)	0 % (0)	35 % (1203)	65 % (2211)
COSMOtherm	66 % (2258)	33 % (1137)	0 % (9)	66 % (2267)	34 % (1147)

490 [a] Φ_G, Φ_W and Φ_{WIOM} represent for fractions of compounds in gas phase, water phase and WIOM phase, respectively.
491 [b] number in brackets are number of compounds

Conclusions

493 For compounds implicated in SOA formation, the prediction of $K_{W/G}$ is much more uncertain

494 than the prediction of $K_{WIOM/G}$. This is true even if we consider that $K_{WIOM/G}$ will vary somewhat

495 depending on the composition of the WIOM (Wang et al., 2015). In particular, the methods

496 currently used for $K_{W/G}$ prediction of these substances have the potential to greatly

497 overestimate $K_{W/G}$. This uncertainty is consequential, as the predicted equilibrium phase

498 distribution in the atmosphere, and therefore also the predicted aerosol yield, is very sensitive

499 to the predicted values of $K_{W/G}$: depending on the method used for prediction, the aqueous

500 phase is either very important for SOA formation from the studied set of compounds or hardly

501 at all. Isaacman-VanWertz et al. (2016) recently found the estimated phase distribution of 2-

502 methylerythritol, an isoprene oxidation product (in Figure S6), highly dependent on the chosen

503 method for predicting $K_{W/G}$. Here we show that this is a general issue potentially affecting a

504 very large number of SOA compounds. In order to identify reliable prediction methods, it will be

505 necessary to experimentally determine the phase distribution of highly functionalized,

506 atmospherically relevant substances, whereby the focus should be on establishing their

507 partitioning into aqueous aerosol.

Data Availability

COSMO files for the 3414 organic compounds can be accessed by contacting the corresponding author.

Supporting Information

The supporting information contains figures and text mentioned in the paper, including detailed information on the organic compounds, e.g. SMILES, molecular formula, molecular weight, functional groups, O:C ratios, predicted K-values, ABSOLV predicted solute descriptors, COSMOtherm predicted vapor pressures and activity coefficients in WIOM and water.

Acknowledgement

We acknowledge funding from Natural Sciences and Engineering Research Council of Canada. Computations for the COSMO files in this study were performed on the General Purpose Cluster (GPC) supercomputer at the SciNet HPC Consortium. SciNet is funded by: the Canada Foundation for Innovation under the auspices of Compute Canada; the Government of Ontario; Ontario Research Fund - Research Excellence; and the University of Toronto.

References

ACD/Labs: Advanced Chemistry Development. Data sheet: Absolv prediction module. Toronto (Canada). [cited 2016 December 2]. Available from: http://www.acdlabs.com/download/docs/datasheets/datasheet_absolv.pdf., 2016.

Arp, H. P. H., Schwarzenbach, R. P., and Goss, K. U.: Ambient gas/particle partitioning. 2: The influence of particle source and temperature on sorption to dry terrestrial aerosols, Environmental Science & Technology, 42, 5951-5957, 10.1021/es703096p, 2008.

Arp, H. P. H., and Goss, K. U.: Ambient Gas/Particle Partitioning. 3. Estimating Partition Coefficients of Apolar, Polar, and Ionizable Organic Compounds by Their Molecular Structure, Environmental Science & Technology, 43, 1923-1929, 10.1021/es8025165, 2009.

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, 10.5194/acp-10-749-2010, 2010.

Compernolle, S., Ceulemans, K., and Muller, J. F.: EVAPORATION: a new vapour pressure estimation method for organic molecules including non-additivity and intramolecular interactions, Atmospheric Chemistry and Physics, 11, 9431-9450, 10.5194/acp-11-9431-2011, 2011.

Endo, S., Pfennigsdorff, A., and Goss, K. U.: Salting-out effect in aqueous NaCl solutions: trends with size and polarity of solute molecules, Environmental Science & Technology, 46, 1496-1503, 10.1021/es203183z, 2012.

Endo, S., Goss, K.-U., Applications of polyparameter linear free energy relationships in environmental chemistry, Environmental Science and Technology, 48, 12477-12491, 2014.

Goss, K.-U.: Prediction of the Temperature Dependency of Henry's Law Constant using Poly-Parameter Linear Free Energy Relationships, Chemosphere, 64, 1369-1374, 2006.

Hallquist, M., Wenger, J. C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N. M., George, C., Goldstein, A. H., Hamilton, J. F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M. E., Jimenez, J. L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T. F., Monod, A., Prevot, A. S. H., Seinfeld, J. H., Surratt, J. D., Szmigielski, R., and Wildt, J.: The formation, properties and impact of secondary organic aerosol: current and emerging issues, Atmospheric Chemistry and Physics, 9, 5155-5236, 2009.

Hilal, S. H., Karickhoff, S. W., and Carreira, L. A.: Prediction of the solubility, activity coefficient and liquid/liquid partition coefficient of organic compounds, Qsar & Combinatorial Science, 23, 709-720, 2004.

Hilal, S. H., Ayyampalayam, S. N., and Carreira, L. A.: Air-Liquid Partition Coefficient for a Diverse Set of Organic Compounds: Henry`s Law Constant in Water and Hexadecane, Environmental Science & Technology, 42, 9231-9236, doi:10.1021/es8005783, 2008.

Hodzic, A., Aumont, B., Knote, C., Lee-Taylor, J., Madronich, S., and Tyndall, G.: Volatility dependence of Henry's law constants of condensable organics: Application to estimate depositional loss of secondary organic aerosols, Geophysical Research Letters, 41, 4795-4804, 10.1002/2014gl060649, 2014.

Ip, H. S. S., Huang, X. H. H., and Yu, J. Z.: Effective Henry's law constants of glyoxal, glyoxylic acid, and glycolic acid, Geophysical Research Letters, 36, 10.1029/2008gl036212, 2009.

Isaacman-VanWertz, G., Yee, L. D., Kreisberg, N. M., Wernis, R., Moss, J. A., Hering, S. V., de Sá, S. S., Martin, S. T., Alexander, M. L., Palm, B. B., Hu, W., Campuzano-Jost, P., Day, D. A., Jimenez, J. L., Riva, M., Surratt, J. D., Viegas, J., Manzi, A., Edgerton, E., Baumann, K., Souza, R., Artaxo, P., and Goldstein, A. H.: Ambient Gas-Particle Partitioning of Tracers for Biogenic Oxidation, Environmental Science & Technology, 50, 9952-9962, 10.1021/acs.est.6b01674, 2016.

Kalberer, M., Paulsen, D., Sax, M., Steinbacher, M., Dommen, J., Prevot, A. S. H., Fisseha, R., Weingartner, E., Frankevich, V., Zenobi, R., and Baltensperger, U.: Identification of polymers as major components of atmospheric organic aerosols, Science, 303, 1659-1662, 10.1126/science.1092185, 2004.

Klamt, A., and Eckert, F.: COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, Fluid Phase Equilibria, 172, 43-72, 2000.

Klamt, A.: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, Elsevier, Amsterdam, 2005.

Kroll, J. H., Donahue, N. M., Jimenez, J. L., Kessler, S. H., Canagaratna, M. R., Wilson, K. R., Altieri, K. E., Mazzoleni, L. R., Wozniak, A. S., Bluhm, H., Mysak, E. R., Smith, J. D., Kolb, C. E., and Worsnop, D. R.: Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nature Chemistry, 3, 133-139, http://www.nature.com/nchem/journal/v3/n2/abs/nchem.948.html#supplementary-information, 2011.

Kurtén, T., Tiusanen, K., Roldin, P., Rissanen, M., Luy, J.-N., Boy, M., Ehn, M., and Donahue, N.: α-Pinene autoxidation products may not have extremely low saturation vapor pressures despite high O:C ratios, The Journal of Physical Chemistry A, 120, 2569-2582, 10.1021/acs.jpca.6b02196, 2016.

Loken, C., Gruner, D., Groer, L., Peltier, R., Bunn, N., Craig, M., Henriques, T., Dempsey, J., Yu, C.-H., Chen, J., Dursi, L. J., Chong, J., Northrup, S., Pinto, J., Knecht, N., and Van Zon, R.: SciNet: Lessons Learned from Building a Power-efficient Top-20 System and Data Centre, Journal of Physics: Conference Series, 256, 012026, 2010.

586   McFiggans, G., Topping, D. O., and Barley, M. H.: The sensitivity of secondary organic aerosol component
587   partitioning to the predictions of component properties - Part 1: A systematic evaluation of some
588   available estimation techniques, Atmospheric Chemistry and Physics, 10, 10255-10272, 10.5194/acp-10-
589   10255-2010, 2010.
590   Mouchel-Vallon, C., Bräuer, P., Camredon, M., Valorso, R., Madronich, S., Herrmann, H., and Aumont, B.:
591   Explicit modeling of volatile organic compounds partitioning in the atmospheric aqueous phase, Atmos.
592   Chem. Phys., 13, 1023-1037, 10.5194/acp-13-1023-2013, 2013.
593   Niederer, C., and Goss, K.-U.: Effect of ortho-chlorine substitution on the partition behavior of
594   chlorophenols, Chemosphere, 71, 697-702, 2008.
595   Platts, J. A., Butina, D., Abraham, M. H., and Hersey, A.: Estimation of Molecular Linear Free Energy
596   Relation Descriptors Using a Group Contribution Approach, J. Chem. Inf. Comput. Sci., 39, 835-845, 1999.
597   Raventos-Duran, T., Camredon, M., Valorso, R., Mouchel-Vallon, C., and Aumont, B.: Structure-activity
598   relationships to estimate the effective Henry's law constants of organics of atmospheric interest,
599   Atmospheric Chemistry and Physics, 10, 7643-7654, 10.5194/acp-10-7643-2010, 2010.
600   Sander, R.: Compilation of Henry's law constants (version 4.0) for water as solvent, Atmospheric
601   Chemistry and Physics, 15, 4399-4981, 10.5194/acp-15-4399-2015, 2015.
602   Schröder, B., Fulem, M., and Martins, M. A. R.: Vapor pressure predictions of multi-functional oxygen-
603   containing organic compounds with COSMO-RS, Atmospheric Environment, 133, 135-144,
604   http://dx.doi.org/10.1016/j.atmosenv.2016.03.036, 2016.
605   US EPA: Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11 or insert version used].
606   United States Environmental Protection Agency, Washington, DC, USA., 2012.
607   Valorso, R., Aumont, B., Camredon, M., Raventos-Duran, T., Mouchel-Vallon, C., Ng, N. L., Seinfeld, J. H.,
608   Lee-Taylor, J., and Madronich, S.: Explicit modelling of SOA formation from alpha-pinene photooxidation:
609   sensitivity to vapour pressure estimation, Atmospheric Chemistry and Physics, 11, 6895-6910,
610   10.5194/acp-11-6895-2011, 2011.
611   Wang, C., Lei, Y. D., Endo, S., and Wania, F.: Measuring and modeling the salting-out effect in
612   ammonium sulfate solutions, Environmental Science & Technology, 48, 13238-13245,
613   10.1021/es5035602, 2014.
614   Wang, C., Goss, K.-U., Lei, Y. D., Abbatt, J. P. D., and Wania, F.: Calculating equilibrium phase distribution
615   during the formation of secondary organic aerosol using COSMOtherm, Environmental Science &
616   Technology, 49, 8585-8594, 10.1021/acs.est.5b01584, 2015.
617   Wang, C., Lei, Y. D., and Wania, F.: Effect of Sodium Sulfate, Ammonium Chloride, Ammonium Nitrate,
618   and Salt Mixtures on Aqueous Phase Partitioning of Organic Compounds, Environmental Science &
619   Technology, 50, 12742-12749, 10.1021/acs.est.6b03525, 2016.
620   Wania, F., Lei, Y. D., Wang, C., Abbatt, J. P. D., and Goss, K. U.: Novel methods for predicting gas–particle
621   partitioning during the formation of secondary organic aerosol, Atmospheric Chemistry and Physics, 14,
622   13189-13204, 10.5194/acp-14-13189-2014, 2014.
623   Wania, F., Lei, Y. D., Wang, C., Abbatt, J. P. D., and Goss, K. U.: Using the chemical equilibrium
624   partitioning space to explore factors influencing the phase distribution of compounds involved in
625   secondary organic aerosol formation, Atmospheric Chemistry and Physics, 15, 3395-3412, 10.5194/acp-
626   15-3395-2015, 2015.
627   Waxman, E. M., Elm, J., Kurtén, T., Mikkelsen, K. V., Ziemann, P. J., and Volkamer, R.: Glyoxal and
628   methylglyoxal setschenow salting constants in sulfate, nitrate, and chloride solutions: measurements
629   and Gibbs energies, Environmental Science & Technology, 49, 11500-11508, 10.1021/acs.est.5b02782,
630   2015.