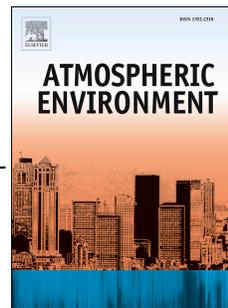


Accepted Manuscript

Modelling the hygroscopic growth factors of aerosol material containing a large water-soluble organic fraction, collected at the Storm Peak Laboratory

Simon L. Clegg, Lynn R. Mazzoleni, Vera Samburova, Nathan F. Taylor, Don R. Collins, Simeon K. Schum, A. Gannet Hallar



PII: S1352-2310(19)30373-5

DOI: <https://doi.org/10.1016/j.atmosenv.2019.05.068>

Reference: AEA 16760

To appear in: *Atmospheric Environment*

Received Date: 14 October 2018

Revised Date: 27 May 2019

Accepted Date: 29 May 2019

Please cite this article as: Clegg, S.L., Mazzoleni, L.R., Samburova, V., Taylor, N.F., Collins, D.R., Schum, S.K., Hallar, A.G., Modelling the hygroscopic growth factors of aerosol material containing a large water-soluble organic fraction, collected at the Storm Peak Laboratory, *Atmospheric Environment* (2019), doi: <https://doi.org/10.1016/j.atmosenv.2019.05.068>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Modelling the hygroscopic growth factors of aerosol material containing a large
2 water-soluble organic fraction, collected at the Storm Peak Laboratory

3 Simon L. Clegg^{a,b}, Lynn R. Mazzoleni^{c,d}, Vera Samburova^e, Nathan F. Taylor^f, Don R. Collins^{f,g}, Simeon K.
4 Schum^d, and A. Gannet Hallar^{e,h}

5 ^a School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

6 ^b Air Quality Research Center, University of California, Davis, California, USA

7 ^c Atmospheric Science Program, Michigan Technological University, Houghton, Michigan, US

8 ^d Department of Chemistry, Michigan Technological University, Houghton, Michigan, USA

9 ^e Division of Atmospheric Sciences, Desert Research Institute, Reno, Nevada, USA

10 ^f Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA

11 ^g Department of Chemical and Environmental Engineering, University of California, Riverside, California,
12 USA

13 ^h Department of Atmospheric Science, University of Utah, Salt Lake City, Utah, USA

14

15 **ABSTRACT**

16 The compositions of six aggregated aerosol samples from the Storm Peak site have been comprehensively
17 analysed (Hallar et al., 2013), focusing particularly on the large water-extractable organic fraction which
18 consists of both high molecular weight organic compounds and a range of acids and sugar-alcohols. The
19 contribution of the soluble organic fraction of atmospheric aerosols to their hygroscopicity is hard to
20 quantify, largely because of the lack of a detailed knowledge of both composition and the thermodynamic
21 properties of the functionally complex compounds and structures the fraction contains. In this work we: (i)
22 develop a means of predicting the relative solubility of the compounds in the water-extractable organic
23 material from the Storm Peak site, based upon what is known about their chemical composition; (ii) derive
24 the probable soluble organic fraction from comparisons of model predictions with the measured
25 hygroscopicity; (iii) test a model of the water uptake of the total aerosol (inorganic plus total water-
26 extractable organic compounds). Using a novel UNIFAC-based method, different assignments of functional
27 groups to the high molecular weight water soluble organic compounds (WSOC) were explored, together
28 with their effects on calculated hygroscopic growth factors, constrained by the known molecular formulae
29 and the double bond equivalents associated with each molecule. The possible group compositions were
30 compared with the results of ultrahigh resolution mass spectrometry measurements of the organic material,
31 which suggest large numbers of alcohol (–OH) and acid (–COOH) groups. A hygroscopicity index (*HI*) was
32 developed. The measured hygroscopic growth is found to be consistent with a dissolution of the WSOC
33 material that varies approximately linearly with *RH*, such that the dissolved fraction is about 0.45 to 0.85 at
34 90% relative humidity when ordering by *HI*, depending on the assumptions made. This relationship, if it
35 also applies to other types of organic aerosol material, provides a simple approach to calculating both water
36 uptake and CCN activity (and the κ parameter for hygroscopic growth). The hygroscopicity of the total
37 aerosol was modelled using a modified Zdanovskii-Stokes-Robinson approach as the sum of that of the
38 three analysed fractions: inorganic ions (predicted), individual organic acids and "sugar alcohols"
39 (predicted), and the high molecular weight WSOC fraction (measured). The calculated growth factors
40 broadly agree with the measurements, and validate the approach taken. The insights into the dissolution of
41 the organic material seem likely to apply to other largely biogenic aerosols from similar remote locations.

42 **Keywords:** Hygroscopicity, soluble aerosol material, organic aerosol composition, aerosol growth factors,
43 thermodynamic modelling.

44

45 **1. Introduction**

46 The hygroscopicity of the mixtures of soluble compounds present in atmospheric aerosols varies in a
47 complex way with aerosol composition, ambient relative humidity (*RH*), and temperature (e.g., Seinfeld and
48 Pandis, 2006; Jacobson, 1999). Water uptake, leading to changes in aerosol size, is a major influence on
49 aerosol optical properties with implications for climate effects, visibility, and health (Boucher et al., 2013;
50 2013; Seinfeld and Pandis, 2006; Pope and Dockery, 2006; Vu et al., 2015). Our ability to quantitatively
51 model hygroscopicity is limited by both practical and theory-related factors: First, our knowledge of aerosol
52 composition is often limited. This is particularly true of the secondary organic fraction (e.g., Kanakidou et
53 al., 2005), but even the analysis of inorganic ions may be incomplete (for example, some ions may not be
54 analysed for, and aerosol acidity cannot be measured directly). Second, the prediction of the equilibrium
55 water activity of aqueous solutions requires complex models (e.g., Fountoukis and Nenes, 2007; Tong et al.,
56 2008, Zaveri et al., 2008; Zuend et al., 2008; Wexler and Clegg, 2002) and remains problematic in low
57 relative humidity conditions for all but the simplest aerosols. The phase state of the aerosol (for example
58 which inorganic components are present largely as solids) may deviate from thermodynamic equilibrium by
59 being supersaturated with respect to one or more salts (Martin, 2000) to an extent that is difficult to
60 determine directly. There are few thermodynamic models that are suitable for predicting the water uptake of
61 the soluble organic fraction of the aerosol, which is known to be complex and contain compounds of widely
62 varying molar mass, functional group composition, and degree of oligomerisation. The UNIFAC model
63 (Fredenslund et al., 1975) was developed for mixtures of water and organic compounds of arbitrary
64 functional group composition, but is primarily intended for molecules containing small numbers of
65 functional groups that are much simpler than the complex structures found in secondary organic aerosols
66 (Hallquist et al., 2009). Nonetheless, it has been adapted and extended to include inorganic ions by Zuend et
67 al. (2008), for use in atmospheric science research, and also incorporated into the Extended Inorganics
68 Aerosol Model (*E-AIM*) of Wexler and Clegg (2002) using the approach demonstrated by Clegg et al.
69 (2001).

70 In the light of the above, simplified treatments of aerosol hygroscopicity are needed, although they should
71 be based upon the measured composition of the atmospheric aerosol to the extent possible (or required by
72 the application). For example, the Zdanovskii-Stokes-Robinson relationship (ZSR) (Stokes and Robinson,
73 1966) has long been used to estimate the water uptake of aerosols in terms of the sums of the amounts of
74 water that would be taken by individual components at the same *RH* and temperature (e.g., Jacobson, 1999;
75 Tong et al., 2008). This relationship is typically applied using individual inorganic salts as components, but
76 is readily extended to treat the total inorganic and organic fractions as components, and using separate
77 models to estimate the water uptake of the two fractions (see section 4 of Clegg and Seinfeld, 2006a). The
78 "kappa" (κ) single parameter representation of aerosol water uptake (Petters and Kreidenweis, 2007), and
79 the derived relationship between aerosol dry diameter and cloud condensation nucleus activity, has also
80 proven very successful in interpreting the results of laboratory measurements of CCN activity. This is true
81 of both simple inorganic aerosols, and complex real aerosols collected in field campaigns (whose
82 composition may not be known). Recently Petters et al. (2017) have explored how the number and location
83 of organic functional groups affect the CCN activity of individual organic compounds.

84 Hallar et al. (2013), and references therein, have presented a detailed chemical analysis of composite
85 aerosol samples collected using a high-volume sampler at Storm Peak Laboratory in Colorado,
86 encompassing inorganic ions, many individual organic acids and sugar-alcohols, and other higher
87 molecular weight water-soluble organic carbon (WSOC, consisting of thousands of individual structures or
88 compounds). Hygroscopic growth factors of the soluble aerosol material, both total and WSOC-only, have
89 been measured using a tandem differential mobility analyser (TDMA) by Taylor et al. (2017), who
90 interpreted their results in terms of the κ parameter (e.g., see their Figure 4). The well characterised

91 composition of the aerosol, especially the organic fraction, allows composition-based approaches to
92 modelling hygroscopicity to be investigated. In this work, we first explore approaches to modelling the
93 water uptake of the WSOC material using UNIFAC and investigate different methods of functional group
94 assignment to the WSOC molecules. The degree to which they are soluble is assessed, at different relative
95 humidities, by comparisons with measured growth factors. Finally, a ZSR-based method is used to estimate
96 the water uptake of the total aerosol as the sum of that calculated for the inorganic fraction (using the *E-*
97 *AIM* model of Clegg and co-workers, Wexler and Clegg, 2002), the organic acids and sugar-alcohols (using
98 UNIFAC), and the separately measured growth factors of the high molecular weight WSOC material.

99

100 2. Data

101 The sampling of atmospheric aerosols was performed from 24 June to 28 July 2010 at the Storm Peak
102 Laboratory (3210 m above sea level, ASL), which is a remote continental site near Steamboat Springs
103 (Colorado, USA). These aerosols are likely to be typical of many remote locations dominated by biogenic
104 aerosol formation and the results of this study are likely to apply to similar aerosols elsewhere. The
105 sampling protocol, treatment of the aerosol samples, and the chemical analyses are described by Hallar et al.
106 (2013). Briefly, the samples were collected on two types of filters: (i) Teflon impregnated glass fiber filters
107 (TIGF, filter size 8" × 10", Fibrefilm T60A20, PALL, Port Washington, NY) for sampling of PM_{2.5} at a
108 flow rate of ~ 1 m³ min⁻¹, and (ii) pre-fired quartz-fiber filters (filter size: 47 mm; 2500 Pallflex QAT-UP,
109 PALL, Port Washington, NY) for sampling of aerosols at a flow rate of ~0.11 m³ min⁻¹. Daily filter samples
110 were combined into six composites (the S1 to S6 that are the subject of this study) based on meteorological
111 conditions and backward trajectories (Hallar et al., 2013). Material from the TIGF filters was used for the
112 analysis of inorganic ions, individual water-soluble organic compounds (Samburova et al., 2013), molecular
113 formula characterization of the higher molecular weight water-soluble organic fraction (Mazzoleni et al.,
114 2012), and hygroscopicity measurements of water extracts with a TDMA (Taylor et al., 2017). These
115 hygroscopicities include those of both the total water-soluble aerosol material (containing the inorganic
116 ions), and measurements for the high molecular weight water soluble organic matter only. Quartz fiber
117 filters were used for analysis of bulk elemental carbon (EC), organic carbon (OC), and water-soluble
118 organic carbon (WSOC) (Hallar et al., 2013; Samburova et al., 2013).

119 All samples were analysed for the inorganic ions Na⁺, K⁺, Mg²⁺, Ca²⁺, SO₄²⁻, NO₃⁻ and Cl⁻ by ion
120 chromatography and automated colourimetry (NH₄⁺ only) (Samburova et al., 2013); and for individual polar
121 organic species (acids, sugars, sugar alcohols, sugar anhydrates, and lignin derivatives) by a combination of
122 IC and GC-MS (Samburova et al., 2013). The extraction of the other water-soluble organic matter (WSOC),
123 using XAD-8 and XAD-4 resins, is also described by Samburova et al. (2013). The WSOC compounds
124 were characterised by ultrahigh resolution Fourier transform-ion cyclotron resonance MS (FT-ICR MS)
125 (see Mazzoleni et al., 2012). The composite samples are referred to in this work as S1-S6 (total water-
126 soluble aerosol material), and SX1-SX6 (water-soluble high molecular weight organic matter only). A
127 schematic diagram of the chemical analyses carried out on the samples is shown in Figure 1 of Hallar et al.
128 (2013).

129 Because the classes of individual polar organic species listed above are only weakly retained by the two
130 resins used to extract the WSOC from the total water-soluble aerosol material, the WSOC extracts contain
131 almost entirely the compounds analysed by FT-ICR MS. Thus only small residual amounts of the ions and
132 individual organic species were analysed in the samples of total water-soluble organic material (Hallar et al.,
133 2013). The results of the FT-ICR MS measurements are expressed in terms of relative amounts of each of
134 the identified molecular formulae, of which there were several thousand for each sample. The amounts of

135 WSOC in the total samples (i.e., before extraction with the resins), and from them the amounts in the
136 aerosol in ng C m^{-3} , were determined by Shimadzu total organic carbon analyser (model TOC-VCSH)
137 (Hallar et al., 2013). The absolute amounts of each of these molecules, per m^3 , were estimated by
138 subtracting the summed concentration of polar organic molecules (in ng C m^{-3}) from the concentration of
139 total water-soluble organic matter to obtain that attributable to the components identified by FT-ICR MS.

140 The relative compositions of composite samples S1-6 are summarised in Figure 1(a,c) as both mass % and
141 mole % of inorganic ions, polar organic molecules, and the higher molecular weight WSOC fraction. It is
142 clear that the inorganic ions dominate the composition of samples S1 and, to a lesser extent, S3. By contrast
143 the ions appear to make up only about 25 mole % of sample S2, and occur in the lowest absolute
144 concentration in sample S5 (3.66 nmol m^{-3} , which is about a factor of 3 lower than in the other samples, see
145 Table 1). The polar organic molecules range from about 11 to 23 mole % of the total sample. Sample S2
146 stands out as containing the largest fraction of WSOC material.

147 The relative compositions of the water soluble organic matter extracts (SX1-SX6) shown in Figure 1(b,d)
148 confirm the low concentrations of both the residual polar organic molecules and inorganic ions. The mole
149 percentage of the WSOC compounds analysed by FT-ICR MS is always greater than 75 mole %, and for
150 two samples is greater than 90 mole %. The mole percentages of ions – which have a large influence on
151 hygroscopicity compared to organic compounds – are well below 10 %. Thus, the measured hygroscopicity
152 of these sample extracts is expected to be controlled mainly by the WSOC compounds (and not the ions or
153 the individual polar organic compounds).

154 We note that growth factors of aerosols of extract SX5 were not measured, thus we have not modelled this
155 property for either SX5 or S5. The compositions of the samples and extracts, in terms of each of the three
156 components, are discussed in more detail below.

157 2.1. Inorganic ions

158 The inorganic compositions of all samples and extracts are listed in Table 1. Total ion concentrations range
159 from 3.66 nmol m^{-3} (S5) to $14.07 \text{ nmol m}^{-3}$ (S1). The dominant anion is SO_4^{2-} in all samples except S4 in
160 which NO_3^- is the principal anion. Ammonium (NH_4^+) is the major cation, followed by K^+ . There are also
161 significant concentrations of Ca^{2+} and Mg^{2+} . The charge imbalances between the cations and anions in each
162 sample – also listed in Table 1 – are large and negative in four out of the six samples. The effect of aerosol
163 H^+ in the samples, if the total SO_4^{2-} in the samples was present as HSO_4^- or $\text{H}_{0.5}\text{SO}_4^{1.5-}$ rather than simply
164 SO_4^{2-} , is shown in the last two rows in Table 1. These alternative charge balances, if realistic, suggest that
165 the levels of acidity in the aerosols are at or between these limits for samples S1-S3, and S6. Sample S5
166 appears to be nearly neutral. Sample S4 is an outlier in these calculations, having apparently low SO_4^{2-}
167 concentration but high NH_4^+ . There is a large excess of positive charge for this sample. The analytical
168 uncertainties in the measured inorganic ion concentrations listed in the Table are of the order of 13%, and
169 do not explain the magnitude of the differences observed. Nor do the results of these comparisons relate in
170 any obvious way to the source trajectories of the samples.

171 The ion concentrations in the extracts SX1-SX6 are lower than in the total sample by well over an order of
172 magnitude, as expected. The magnitude of the charge imbalances (see last line of Table 2), which for these
173 samples are mostly positive, are likely to be due to the effects of the larger experimental uncertainty in the
174 determination of ion concentrations. However, because the ions have very low absolute concentrations in
175 the extracts (see Fig. 1d) the effect of errors on the calculated hygroscopicity will be small.

176 2.2 Polar organic compounds

177 The concentrations of the 47 polar organic compounds measured by Samburova et al. (2013) are
178 summarised in Table 3 and listed individually in Table S1 of the Supplementary Information. For full
179 details, see Tables S3-S5 in the Supplementary Information to Samburova et al. (2013). The total
180 concentrations range from 0.9 (S5) to 2.29 nmol m⁻³ (S1), with the bulk of the compounds consisting of low
181 molar mass acids, and sugars.

182 2.3 WSOC compounds

183 The organic compounds determined by Mazzoleni et al. (2012) in the water-soluble organic material
184 extracts using ultrahigh resolution FT-ICR MS are summarised in Table 4 (and are listed in full in the
185 Supplementary Information to their publication). Total concentrations, per m³, are given for both the S1-6
186 and SX1-6 samples. Very large numbers of molecules (containing two or more of C, H, O, N, and S atoms)
187 were determined – 3881 in the case of sample S4 – a large fraction of which are common to all samples, as
188 shown by the last line in Table 4. Thus, the WSOC organic material in the six composite samples appears to
189 be rather similar. The mean, concentration weighted, number of carbon atoms in each molecule is about 17
190 in all samples, and the mean molar masses vary over a relatively small interval (368.5 to 392.1 g). The
191 numbers of carbon atoms in the molecules range from 3 to 45. Thus, it would be expected that some
192 fraction of this organic material (i.e., the molecules with large numbers of carbon atoms) might be insoluble
193 in water at the relatively high liquid phase concentrations encountered during the hygroscopicity
194 measurements. Relative abundances of molecules containing different numbers of carbon atoms are
195 discussed by Mazzoleni et al. (2012), and illustrated in their Figures 4 to 6.

196 In addition to the full scan analysis, FT-ICR MS/MS fragmentation analysis was used to investigate the
197 functional groups present in the identified molecular formulae in sample extract SX4. Due to the extreme
198 isobaric complexity of water-soluble organic aerosol, individual mass spectral peaks could not be isolated for
199 fragmentation. Instead small mass range windows (6 or 10 u) were selected for fragmentation, consistent with
200 LeClair et al. (2012). Each mass window was defined by a central mass selected at intervals of 5 u (m/z 180,
201 185, etc.) over the range of m/z 160 - 365 and then every 10 u over the range of m/z 365-485. Ultrahigh
202 resolution analysis using FT-ICR MS was done on both unfragmented ions (representing precursor ions) and
203 the fragmented ions after collision induced dissociation (representing product ions). Molecular formulae were
204 then assigned to the collected ultrahigh resolution mass spectra using Composer software as described in
205 Mazzoleni et al. (2012). The resulting precursor and fragment formulas were paired based on the exact mass
206 differences associated with expected common neutral losses (e.g., CO₂, H₂O, etc.). A total of 1471 precursor
207 formulas were assigned to the studied mass ranges and 100% of them were also found in the full scan analysis
208 of this same sample reported in Mazzoleni et al. (2012).

209 Quantitative information regarding the molecules in this component of the aerosol is limited to the amounts,
210 numbers of C, H, O, N, and S atoms in each molecule, and the numbers of double bond equivalents (DBE).
211 These are defined by: $DBE = C - H/2 + N/2 + 1$, where C, H, and N are the numbers of atoms of each of the
212 three elements in the molecule. Note that one ring counts as 1 DBE, a triple bond counts as 2 DBE and an
213 aromatic ring is 4 DBE (one for the ring plus one for each C=C). It is polar groups such as –OH and –
214 COOH that particularly influence solubility in water and the relationship between water activity
215 (equilibrium *RH*) and concentration or hygroscopicity. The results of the FT-ICR MS analysis give some
216 insight into the abundances of the different functional groups in the WSOC material for sample SX-4, in the
217 following way. A total of 21 different neutral losses were observed for the studied precursor and fragment
218 ion molecular formulas based on exact mass difference pairing. For example, a neutral loss of H₂O is
219 indicative of a hydroxyl functional group (–OH) and a neutral loss of CO₂ is indicative of a carboxyl
220 functional group (–COOH). Likewise, a neutral loss of CH₂O₃ is indicative of two functional groups
221 (carboxyl (–COOH) and hydroxyl). Combinations of neutral losses are expected for multifunctional

222 compounds such as those present in water-soluble organic aerosol, and multiple neutral losses were found
223 to be associated with many of the precursor formulae. Roughly 70% of the precursor formulae showed 5 or
224 more neutral losses and 36% showed 8 or more neutral losses. This high number of neutral losses suggests
225 the presence of multiple structural isomers per assigned molecular formula, an observation supported by
226 Zark et al. (2017).

227 The neutral losses were grouped into five major categories: CO₂ losses, H₂O losses, methoxy losses,
228 aldehyde losses, and nitrogen and/or sulfur losses. Some neutral losses can fit into two categories, for
229 example the CH₂O₃ loss mentioned previously is classified as both a CO₂ loss and an H₂O loss because both
230 functional groups are contained within that neutral loss. This means that some losses will be counted twice,
231 once in two different categories. The two most abundant loss categories were CO₂ and H₂O, which were
232 observed for 1279 (86.9%) and 1339 (91.0%) of the precursor formulas overall. The two next most
233 abundant were aldehyde (1148, 78.0%) and methoxy (978, 66.5%) group neutral losses. The complete
234 breakdown of this is shown in Table 5, and the results are discussed further in the Appendix.

235

236 3. Functional Group Compositions and Hygroscopicity of the WSOC Compounds

237 Predictions of the water uptake of the WSOC fraction of the aerosol, apart from the simple assumption of
238 Raoult's law behaviour and the fraction of the material that dissolves, require estimates of the compositions
239 of the individual molecules in terms of the functional groups present. With this knowledge, thermodynamic
240 models such as UNIFAC (Fredenslund et al., 1975) can be used. In this work, we compare both approaches.

241 The UNIFAC model predicts the activities of the constituents of liquid mixtures of organic compounds and
242 water based upon the compositions of the molecules in terms of their functional groups. The model contains
243 parameters that express the interactions between the functional groups, which have been determined by
244 fitting vapour/liquid equilibrium and other data for very large numbers of liquid mixtures. Neither the
245 positions of the groups within each molecule, nor the effects of scaling when multiple instances of a single
246 group are present in a molecule (their contributions are broadly additive), are considered. The defined
247 functional groups are restricted to those for which there are data. These are mostly from measurements for
248 compounds that are used in industry and/or are common in nature, and which generally contain few
249 functional groups. This contrasts with the composition of organic aerosols which analysis has shown to be
250 multifunctional, have quite complex structures, and contain groups which are not currently included in
251 UNIFAC.

252 Because of the above limitations, estimates of the properties of this component of the organic fraction of the
253 aerosol using UNIFAC can only be considered approximate at best. The calculations in this work, using
254 UNIFAC, are probably best viewed as best semi-quantitative estimates of how the effects of non-ideality
255 might affect calculated hygroscopicity relative to the assumption of Raoult's law. The groups in the
256 UNIFAC model employed in this study are those listed by Hansen et al. (1991), Wittig et al. (2003), and
257 Balslev and Abildskov (2002). Although these groups may only represent a subset of those present in the
258 molecules in the samples, it is also true that the groups with the greatest influence on hygroscopicity are
259 likely to be the highly polar ones which are well represented in the model. Sulphur and nitrogen containing
260 groups, of which there are few in UNIFAC, are likely to be of little importance because the molecules
261 containing them are present at very low concentrations in this fraction of the aerosol (an average of 0.23
262 assigned sulphone, sulphide, thiol, or "nitro" groups per molecule, where the average total number of
263 groups is 12.3).

264 We have estimated the compositions of the high molecular weight WSOC material in the samples in terms
 265 of the UNIFAC functional groups, based on the following assumptions: (i) the molecules consist either of
 266 chains of carbon atoms (with branches, if necessary, but not aliphatic rings), or a single aromatic ring with
 267 either one or two carbon chains attached; (ii) the only functional and structural groups present are those
 268 available within UNIFAC. This work, described in the Appendix, has enabled the calculation of equilibrium
 269 *RH* as a function of concentration for the WSOC compounds, including the effects of non-ideality, and also
 270 the development of a hygroscopicity index (see below) to help account for their solubilities.

271 3.1 Hygroscopicity index

272 The molecules in the WSOC material contain from 3 to 45 C atoms each. This large range in the number of
 273 C atoms implies significant variations in their solubility in water, and consequently hygroscopicity. Indeed,
 274 the molecules with the most carbon atoms seem unlikely to be soluble, even if they also contain polar
 275 functional groups. Also, the more aliphatic molecules – which tend to be larger – are less likely to be
 276 miscible with water. In both cases, low solubility and low miscibility, the molecules will not contribute
 277 significantly to hygroscopicity. We have constructed a hygroscopicity index to attempt to assess this
 278 behaviour in a semi-quantitative way. We define the index value, $HI_{(i)}$, of a WSOC compound i by:

$$279 \quad HI_{(i)} = \log_{10}(x_i^* f_i^\infty) \quad (1)$$

280 where f_i^∞ is the activity coefficient of organic compound i at infinite dilution in water, relative to a
 281 reference state of pure liquid i , and x_i^* is the dry mole fraction of the compound in the WSOC sample of
 282 interest (i.e., not including water in the denominator). Values of f_i^∞ are calculated with UNIFAC, using the
 283 estimated functional group compositions from the Appendix, and are higher the less miscible in water the
 284 compound. The inclusion of x_i^* in the index takes account of the differing amounts of the compounds
 285 present: a largely non-miscible or insoluble compound may dissolve in water and contribute to the
 286 hygroscopicity if its concentration is very low (and therefore very dilute in the solution).

287 We have calculated HI values for all WSOC molecules in the six sample extracts, and in Figure 2a they are
 288 shown plotted against cumulative mole fraction for sample SX1. The compounds are ranked in order of
 289 increasing HI . These calculations are for the base case group assignments (minimising the number of
 290 UNIFAC groups needed to describe each molecule). Results for the other samples are similar. Recall that,
 291 for a calculated activity of an organic compound in water ($x_i f_i$), where x_i is the mole fraction of compound
 292 i in solution, a value of unity indicates a concentration beyond which no dissolution of the compound can
 293 occur. Further additions of the compound, if it is liquid at the temperature of interest, would result in a
 294 phase separation. If it is a solid, then precipitation of the solid from solution would presumably have
 295 occurred at some lower activity. High values of HI correspond to non-hygroscopic, and probably insoluble
 296 compounds (even at high *RH*), while compounds with lower values are expected to be more soluble and
 297 hygroscopic over a wide *RH* range. A value of HI equal to unity does not have any particular significance.

298 The overall shape of the curve in Figure 2a suggests that there are relatively few compounds – less than
 299 about 25 mol% of the total material in the sample – with low solubility, and most of the compounds occupy
 300 a broad intermediate range. At low equilibrium *RH* it is expected that only the most soluble compounds (at
 301 the left of the plot) will dissolve and contribute to hygroscopicity. At higher *RH*, where more water would
 302 be present in the aerosol and the mole fractions (x_i) of the organic compounds lower, a greater fraction of
 303 the compounds would be expected to dissolve. How do values of the index relate to carbon number and to
 304 the O:C ratios of the molecules? Both quantities are plotted for sample SX1 in Figure 2b, again ranked in
 305 order of increasing HI . The expected relationships, that low carbon number and high O:C ratio corresponds
 306 to high miscibility and hygroscopicity (and high carbon number and low O:C to low miscibility) can be
 307 seen in the figure. However, for the bulk of the material, of the order of 75%, there is considerable scatter

308 and the relationships are approximate only. In part this reflects the varying amounts present of compounds
309 that may have similar carbon numbers and O:C ratios (and perhaps f_i^∞), but very different x_i^* .

310 3.2 Categorising the compounds

311 In order to investigate the variation of functional group composition with HI for sample SX1 we have
312 divided the material shown in Figure 2 into 5 fractions containing equal moles of material: 0-20%, 20-40%,
313 etc., so that the first group contains the most soluble fraction of the WSOC material (lowest HI values) and
314 the fifth and last group contains the least miscible or soluble material (highest HI values). The average
315 formulae of each group are shown in Figure 3a. As expected, the numbers of C and H atoms increase going
316 from left to right (soluble to insoluble), although the numbers of O atoms vary little – evidently it is the
317 increasing numbers of C and H that account for the reduction in expected solubility. Figure 3b shows the
318 assigned UNIFAC group compositions of the most soluble ($\log_{10}(HI) < -1.75$), least soluble ($\log_{10}(HI) >$
319 2.25), and intermediate solubility ($0 < \log_{10}(HI) < 1.0$) fractions of the material. The clearest features are,
320 first, the highly aliphatic nature of the least soluble fraction, and its relative simplicity: there is little N and
321 S, and the bulk of the O present is predicted to be in the form of acetate and ether groups. By contrast, the
322 most soluble fraction (see Figure 3c) has only small numbers of alkane and alkene groups, but a variety of
323 the more complex and polar groups dominated by ether and acetate but with also a significant number of –
324 OH and –COOH.

325 It is clear from Figure 3 that the ranking of the compounds by hygroscopicity index is broadly consistent
326 with what is expected: compounds that are largely aliphatic in nature are expected to be insoluble, and
327 therefore unlikely to contribute to hygroscopicity; the more chemically complex and less aliphatic
328 compounds are expected to be most miscible and/or soluble. The hygroscopicity index, and rankings, are
329 essentially qualitative and, in particular, the available functional groups do not represent the full range of
330 those that occur in aerosol organic material. However, the index is helpful in exploring the influence of
331 varying WSOC solubility on the predicted hygroscopicity of the total aerosol material and extracts as will
332 be shown further below.

333 3.3 Varying the assignment of functional groups

334 The effect of alternative functional group assignment methods on the estimated composition of S1
335 molecules is summarised in Table A1. Maximising the number of functional groups per molecule strongly
336 favours the assignment of alkane, alkene, alcohol and aldehyde groups over all others (see the second
337 column of results in the table). Assigning a high weight to alkane, alcohol, and acid functional groups
338 results in a large reduction in the assigned number of aldehyde and alcohol groups relative to the previous
339 case and their replacement by the acid group –COOH (last column in Table A1). Figure 4 summarises the
340 average estimated compositions of the SX1 sample fractions 1 (most hygroscopic), 3, and 5 (least
341 hygroscopic) for these two additional cases. The molecules were grouped into the five fractions according
342 to the calculated HI , in the same way as for the base case. In a comparison of Figure 4 with the base case
343 (Figure 3b,c) several features stand out. First, the compositional simplicity of the molecules for the two
344 additional cases and, second, the much smaller variation in the predicted number of alkane groups per
345 molecule. It appears that variations in predicted degree of hygroscopicity of the molecules is driven mostly
346 by the number of –OH (alcohol) groups for the case where the number of functional groups per molecule is
347 being maximised. Here, the predicted average of about 10 –OH per molecule (Figure 4a) in the most
348 hygroscopic fraction is clearly too high as this would be typical of the sugars and sugar alcohols that were
349 analysed by IC and MS and only present at very low concentration on the WSOC material. For the case
350 where alkane, alcohol, and acid groups were given high weight (Figure 4b) the predicted hygroscopicity is
351 driven partly by the numbers of alkane groups (more of these groups means a lower hygroscopicity) and

352 partly by the combined number of predicted –OH and –COOH groups (which are highest in fraction number
 353 1). Finally, a comparison of the absolute values of the predicted hygroscopicity index for the three cases
 354 suggests that the alternative group assignments yield molecules that are much more likely to be soluble and
 355 hygroscopic. This is due to the large numbers of acid and particularly alcohol functional groups that are
 356 predicted. The effects of the alternative group assignments on calculated aerosol water uptake are examined
 357 in section 5.

358

359 4. Methods

360 In this section we describe the methods used to estimate the hygroscopicity of the aerosol material so that
 361 the results of the modelling can be compared with measured hygroscopic growth factors (*GF*). This quantity
 362 is defined by the following equation:

$$363 \quad GF = [(\text{Volume at the } RH \text{ of interest}) / (\text{Volume at a reference } RH)]^{1/3} \quad (2)$$

364 where the reference relative humidity (*RH*) is 10% in our growth factor determinations. The compositions
 365 of the aerosol samples in this study are known in terms of the concentrations of inorganic ions, polar
 366 organic molecules, and other WSOC molecules. The calculation of the hygroscopic growth factors, to
 367 compare to the measurements, requires that the water content of the aerosols be calculated as a function of
 368 *RH* (equivalent to the water activity, a_w , of the droplets), followed by their densities and hence the total
 369 volume of the aerosol at each concentration.

370 The water content of an aqueous mixture, containing two or more solutes, can be estimated using the
 371 Zdanovskii-Stokes-Robinson (ZSR) relationship (Stokes and Robinson, 1966), so that:

$$372 \quad \sum_i (m_i / m_i^\circ) = 1 \quad (3)$$

373 where m_i is the molality of each solute i in the mixture, and m_i° is the molality of i in a pure (single solute)
 374 solution of i at the water activity of the mixture. This relationship can also be expressed, more simply, as:

$$375 \quad W_T = \sum_i w_i^\circ \quad (4)$$

376 where W_T is the total mass of water in the mixture, and w_i° is the mass of water associated with the moles of
 377 each solute i in a pure (single solute) solution of i at the water activity of the mixture (see equation (7) of
 378 Clegg et al., 2003). Clegg and Seinfeld (2006b) have shown, in their section 7, that the w_i° in the equation
 379 above can also refer to groups of solutes within the overall mixture. We apply this principle here so that, for
 380 each aerosol sample:

$$381 \quad W_T = W^\circ(\text{ions}) + W^\circ(\text{polar organic}) + W^\circ(\text{WSOC}) \quad (5)$$

382 where W° is the mass of water associated with the named group of solutes in a solution containing only
 383 these solutes, at the water activity of the mixture. In equation (5) “ions” refers to the inorganic electrolytes
 384 in the aerosol samples (Table 1), “polar organic” to the polar organic molecules (Table 3), and “other
 385 organic” to the WSOC organic molecules analysed by FT-ICR MS (Table 4).

386 Analogous relationships to ZSR can be derived for other thermodynamic and physical properties, and Hu
 387 (2000) has determined such an equation for the density of solution mixtures (his equation 11), which can be
 388 transformed into an additive relationship for solution volumes. Applied to the system of interest here, it
 389 yields:

$$V_T = V^{\circ}(\text{ions}) + V^{\circ}(\text{polar organic}) + V^{\circ}(\text{WSOC}) \quad (6)$$

where V_T is the total volume of the aqueous mixture at water activity a_w , and V° are the volumes occupied by aqueous solutions of the three named groups of solutes at the water activity of the mixture. We have calculated the volume of each individual mixture in equation (6) using equation (12) of Semmler et al. (2006):

$$1/\rho = \sum_i x_i^* / \rho_i^{\circ} \quad (7)$$

where ρ is the density of the mixture, and ρ_i° is the density of a pure aqueous solution of solute i at the total weight fraction of solutes in the mixture, x_i^* , given by:

$$x_i^* = n_i / \sum_j n_j \quad (8)$$

where n_i is the number of moles of solute i in the mixture, and the summation is over all solutes j . The total volume of each of the three components of the aerosol (the V° in equation 6) is related to its density by $V^{\circ} = M_T / \rho$ where M_T is its total mass. For the polar organic and WSOC components of the solution, the individual solutes i are the organic molecules. The composition of the inorganic component of the solution is expressed in terms of individual electrolyte solutes, rather than ions, using equation (5) of Clegg and Simonson (2001) for x_i^* .

The calculation of the water content and volumes of the three components of the aerosols (ions, polar organic, and WSOC) is described in more detail in the sections below.

4.1 Inorganic electrolytes (ions)

The electrolyte components of the samples of total aerosol material will take up most of the water. It is shown in Table 1 that there are charge imbalances between the cations and anions, and that these are strongly negative for the samples that contain the most sulphate. The last two rows of the table show that the balance is improved by the assumption that the sulphate is present in the aerosol as HSO_4^- , or $\text{H}_{0.5}\text{SO}_4^{1.5-}$ for all samples except S5 (for which the charge balance is in error by only 4%) and S4. The NH_4^+ concentrations in aerosol samples S1-S3, and S6, are consistent with this, implying the presence of ammonium bisulphate or letovicite in the aerosol. However, the significant concentrations of NO_3^- and Cl^- present would not generally be expected in a strongly acidic aerosol because they would be lost to the gas phase as HNO_3 and HCl .

Calculations of the volumes of the measured inorganic components of the aerosol samples, as a function of RH , were carried out for a number of different cases. In the first of these the charge imbalance between cations and anions was corrected by adjusting both the cation and anion amounts so that the total charges ($\sum_c n_c z_c$ and $\sum_a n_a |z_a|$) were both equal to the mean value of the two sums obtained from the measured amounts in Table 1. For the second and third cases it was assumed that either the existing measured cation or anion concentrations were correct, and ions of the other charge type were then adjusted to give charge balance. We also carried out calculations for one acidic case (a negative charge imbalance was corrected by adding H^+).

The solubility of CaSO_4 (as gypsum, $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) is very small (about $0.015 \text{ mol kg}^{-1}$ at 25°C) and all Ca^{2+} present in the aerosol samples was assumed to remain as a solid at all RH . This leaves the ions NH_4^+ , Na^+ , K^+ , Mg^{2+} , SO_4^{2-} , NO_3^- and Cl^- potentially dissolved in the aqueous phase. The *E-AIM* Model III of Clegg et al. (1998) was used to calculate the water uptake of the inorganic ions and the particle volumes. Ion interactions between cations Mg^{2+} and K^+ (which are not present in Model III) and the anions SO_4^{2-} , NO_3^- , and Cl^- were added as described in the Supplementary Information. The densities and volumes of the

431 particle solutions, and the solid salts that form at low *RH*, were calculated using the work of Clegg and
432 Wexler (2011), and equation (7) above. The inclusion of the additional electrolytes and solid salts, for the
433 volume calculations, is also described in the Supplementary Information.

434 4.2 Polar organic compounds

435 Molar volumes of the polar organic compounds summarised in Table 3 were calculated from their molar
436 masses, and (liquid) densities estimated using the method of Girolami (1994). This is based upon the
437 formulae of the compounds and also the numbers of particular chemical groups that are present (notably
438 alcohol, acid, amide, sulphoxide and sulphone). These molar volumes were assumed to be constant for all
439 solution water contents. The method of Girolami is one of several assessed by Barley et al. (2013), and
440 found to yield densities to within 10% of the true value in almost all cases, and 5% in most. A fixed density
441 of 1.3 g cm^{-3} was assumed for all compounds as solids.

442 The water uptake of the aqueous solutions of polar organic compounds was calculated using UNIFAC
443 (described in section 3.1). For simplicity, all of the polar compounds were assumed to be completely
444 soluble at relative humidities above the reference *RH*. Given that these compounds constitute a minor
445 fraction of the total solutes (12 to 25 mol% of the total samples, and only 1.5 to 18 mol% of the extracts),
446 this assumption is unlikely to have a large effect.

447 4.3 WSOC fraction

448 The liquid molar volumes of the WSOC compounds were estimated in the same way as for the polar
449 organic compounds described above and making use of the UNIFAC group assignments in the three
450 different cases being examined. The organic compounds in solid form are treated in the same way as the
451 polar compounds and assumed to have a density of 1.3 g cm^{-3} . The numbers of C atoms in the WSOC
452 molecules range from 3 to 45, suggesting that not all of the higher molecular weight WSOC fraction is
453 soluble (sections 3.3 and 3.4). This is a feature of WSOC behaviour that is explored in comparisons with
454 measured hygroscopic growth factors in the next section.

455 456 5. Modelling Hygroscopic Growth Factors of the Organic Material

457 It is well understood that electrolyte solutes are more hygroscopic, and have higher growth factors, than
458 most soluble organic compounds. This is illustrated in Figure 5, which compares measured growth factors
459 of the SX1 WSOC material, calculated growth factors of the S1 polar organics, and calculated growth
460 factors of $(\text{NH}_4)_2\text{SO}_4$, $(\text{NH}_4)_3\text{H}(\text{SO}_4)_2$, and NH_4HSO_4 . The polar organic fraction was assumed to be fully
461 liquid at all *RH* (hence the continuous increase in GF with *RH*, in contrast to the deliquescence transitions
462 shown for the salts). The growth factor would be increased by a factor of about 1.057 if the polar organic
463 fraction were assumed to be solid at the reference *RH* (this is based upon the difference between the solid
464 and liquid molar volumes of glucose, which is a significant component of the polar organic fraction). The
465 measured growth factors of the WSOC material in Figure 5 are very low, which is also true of the other
466 composite samples (see Figure 5 of Hallar et al., 2013). A simple calculation suggests that, at 80% *RH*,
467 about 840 g of WSOC material is required to take up the same amount of water as 1 mole (132 g) of
468 ammonium sulphate. Figure 1a shows that a comparable mass ratio of WSOC material to inorganic
469 electrolytes (840 : 132, or 6.3 : 1) is approached only in sample S2. In the other samples, the ratio varies
470 from about 1:1 (S1) to about 2.5:1 (S5).

471 Although the WSOC material contributes little to the water uptake of the total aerosol for most samples, its
472 hygroscopicity is still of interest, for three reasons: first, because some reactions involving WSOC

473 compounds may only occur in the aqueous phase, or at an interface between a solid and an aqueous phase
474 (Hallquist et al., 2009; Smith et al., 2014; and references therein). Second, a knowledge of how WSOC
475 material interacts with water is important for understanding the physical state of the aerosol. Our
476 thermodynamic treatment of the WSOC fraction of aerosol as being partially soluble at room temperature
477 corresponds to the "semi-solid" state discussed by Shiraiwa et al. (2017), and the high viscosity semi-solid
478 or glassy secondary organic aerosols examined by Petters et al. (2018). Freedman (2017) discusses the
479 effect of the organic component of the aerosol on particle morphology of aerosol particles. Third, the
480 composition of the original particles is likely to have been much more diverse than that of the aggregate
481 samples, with many having a higher fractional organic content than suggested by the composite average.
482 For these particles the WSOC hygroscopicity may largely control their water uptake and contribution to
483 CCN concentrations in the atmosphere. The variation of particle composition (inorganic vs. organic) with
484 particle size, and effects on the hygroscopicity of aerosols observed during the MILAGRO field study are
485 described by Wang et al. (2010).

486 In the sections below, calculations of the hygroscopicity of each fraction of the aerosol are discussed and
487 compared with the measurements of Taylor et al. (2017).

488 5.1 Polar organic compounds

489 The 48 polar organic compounds for which concentrations were measured individually in all samples by
490 Samburova et al. (2013) are listed in Table S1, and their UNIFAC group assignments are listed in Table S5
491 of the Supplementary Information. These were used in the calculation of the hygroscopicity of this fraction
492 of the aerosol and its contribution to the total volume, and hence growth factor, of the aerosol material. The
493 predicted growth factors of the polar organic compounds in sample S1, relative to a hypothetical liquid
494 mixture at 20% *RH*, are intermediate between those of the ammonium sulphate salts and the measured
495 values of the WSOC material (Figure 5). However, the polar organic compounds account for only 20 ± 11
496 mass % of the total water-soluble organic material (Samburova et al., 2013), or an average of 39 mol %,
497 which suggests that their contribution to the water uptake of the total aerosol material will be modest. The
498 measured values for the total water-extractable aerosol material (Figure 2 of Taylor et al., 2017) show
499 uptake of water by the aerosols at all *RH*. Consequently, in the comparisons made in section 5.3 below, we
500 assumed that the polar organic compounds mix with this water at all *RH* and do not occur as solids in the
501 aerosol.

502 5.2 The WSOC fraction

503 A number of different physical states of this fraction of the aerosol material can be envisaged. These are
504 shown in Figure 6. The first case, in which all molecules are assumed to be fully miscible with water, is the
505 simplest. However, this appears unlikely to be realistic given the large fraction of WSOC material
506 containing molecules with 20 or more carbon atoms. The next possible state, case (2), is one in which
507 aerosol particles consist of a core of insoluble or slightly soluble molecules, surrounded by an aqueous
508 phase containing soluble molecules. The higher the relative humidity, the smaller this core might be (as
509 greater proportions of the less soluble molecules are able to dissolve into the larger volume of aerosol
510 water). The third case is one in which there exists a hydrophobic organic liquid phase, containing very little
511 water and contributing very little to the growth factor, in equilibrium with an aqueous phase containing the
512 more polar, soluble, organic molecules. The final case shown in Figure 6 is a combination of cases (2) and
513 (3): an insoluble or partially soluble core and two liquid phases. It seems reasonable to expect that this last
514 case – if realistic – would yield the lowest growth factors.

515 Freedman (2017) has discussed phase separation, and the existence of more than one liquid phase, in
516 organic aerosols. The "partially engulfed" morphology in Freedman's Figure 6, for a particle consisting of

517 two immiscible liquid phases, is thermodynamically equivalent to cases 3 and 4 in our Figure 6: two
518 immiscible liquid phases in equilibrium (case 3), and the same but with an insoluble core (case 4). By
519 "thermodynamically equivalent" we mean that the equilibrium state of the gas/particle system is the same.
520 However, it is recognised that the geometry and arrangement of the two phases in the particle might affect
521 responses to changes in the surrounding atmosphere: e.g., a particle in which an inner liquid phase core was
522 completely surrounded by a "shell" of a second liquid phase would be expected to come to equilibrium with
523 the surrounding atmosphere more slowly than would our case 3 or the "partially engulfed" case of
524 Freedman (2017).

525 It is not possible to model directly the growth factors for all the cases shown in Figure 6. The true
526 functional group compositions of the molecules are not known, nor are the solubilities of the individual
527 solid compounds in the WSOC material. The occurrence of more than one liquid phase can, in principle, be
528 modelled using UNIFAC. Test calculations for the base case UNIFAC group assignments (minimising the
529 number of functional groups per molecule) did suggest the formation of more than one liquid phase at high
530 *RH*. However, the very large number of molecules combined with the uncertainty as to their true group
531 compositions, and the fact that the model performs relatively poorly for molecules containing multiple polar
532 functional groups, means that the results would be unlikely to be accurate.

533 Case (2) is the simplest of the three cases shown in Figure 6 that do not assume a single liquid phase, and
534 was investigated in the following way. Having first ranked the WSOC molecules in the SX1 extract in order
535 of their hygroscopicity (high to low, by their calculated *HI*), we computed the growth factors assuming that
536 a range of different fractions of the WSOC material could dissolve in water: first, the most soluble 5 mol %,
537 then the most soluble 25 mol %, and so on until it was assumed that 85 mol % of the molecules dissolve in
538 water. The results are compared with the measured growth factors of sample SX1 in Figure 7. The
539 relationship between water content and water activity (hence *RH*) was calculated using UNIFAC in plots (a-
540 c), and Raoult's law in plot (d). The most obvious feature of the figure is that the measured growth factors
541 match the calculated values for increasingly high soluble fractions as *RH* rises. This corresponds to more of
542 the WSOC material dissolving at high *RH*: for example about 45 mol % at 80% *RH*, and 70 mol % at 90 %
543 *RH* in Figure 7(a). This, qualitatively, is what is expected for the assumed physical state of the WSOC
544 aerosol material embodied in case (2) in Figure 6.

545 There are features of the plots in Figure 7 that require explanation. First, the predicted growth factors of the
546 partially dissolved WSOC are greater than unity at 20 % *RH* and increase with the assumed soluble fraction.
547 This is because they are all referenced to a dry aerosol. Thus, for example, Figure 7(a) shows that at 20 %
548 *RH* the SX1 WSOC aerosol in which 85 mol % of the WSOC molecules are soluble has a growth factor of
549 1.07 relative to a dry aerosol at the same *RH*. The growth factors at the lowest *RH* – for which water has the
550 smallest influence – largely reflect the difference between the assumed 'dry' density of the WSOC material,
551 which is 1.3 g cm^{-3} , and those of the liquid organic molecules estimated using the equation of Girolami
552 (1994). Changing the assumed dry density affects the calculated growth factors at all *RH*: a decrease to 1.2
553 g cm^{-3} results in a reduction of the predicted growth factor for 85 mol % soluble material from 1.117 to
554 1.094 at 80% *RH*, and changing it to 1.4 g cm^{-3} increases the calculated growth factor to 1.139 at the same
555 *RH*. Second, the large differences between the calculated growth factors in Figure 7(a-c), even at low *RH*,
556 are caused by differences between the predicted water activities of the aqueous phase, differences in the
557 predicted molar volumes of the molecules (related to their group compositions), and because the molecules
558 are ordered differently (by the calculated *HI*) in each of plots a-c. Thus, for example, the 25% of the
559 molecules in Figure 7(a) predicted to have the highest solubility are not the same as in plots (b) and (c). In
560 the final calculation, for Raoult's law water uptake, we assumed the same ordering of molecules as in plot (a)
561 and the growth factors for the two cases are quite similar at the lowest *RH*.

562 In summary, we can say that: (i) measured growth factors of the WSOC fraction are consistent with a
563 degree of solubility that varies with *RH*, and complete dissolution of the WSOC material is not approached
564 in any of our calculations until at least 90% *RH*; (ii) the dissolved fractions of WSOC material (at a chosen
565 *RH*) that can be inferred from the results in Figure 7 differ according to the assumptions made in each of the
566 four cases. The Raoult's law case in Figure 7(d), which is the simplest to model, yields a larger predicted
567 soluble fraction at moderate *RH* (50-60%) than the other cases, but this isn't true at high *RH*. (iii) Figure 7(c)
568 shows results for the case for which the UNIFAC group assignments are best supported by the FT-ICR MS
569 results (high weight given at alkane, -OH and -COOH groups). The results for both these cases are
570 consistent with the expected greater hygroscopicity of polar organic compounds (in the sense of higher
571 water uptake per amount of soluble material) and also suggest that not all of the organic material dissolves
572 even at the highest *RH*. This seems reasonable given the large number of carbon atoms in many of the
573 molecules.

574 The relationship between soluble fraction and *RH*, and the relevance of our results to other types of organic
575 aerosol material and to aerosol (atmospheric) models, is discussed in section 7.

576

577 6. Modelling Hygroscopic Growth Factors of the Total Soluble Material

578 Because of the difficulties of modelling water uptake of the WSOC fraction, described above, we have
579 calculated the water uptake of the total aerosol as the sum of the measured WSOC uptake, and the predicted
580 water uptake of the polar organic compounds and the inorganic ions. The calculation of water uptake of the
581 inorganic ion fraction of the aerosol ($W^o(\text{ions})$ in eq. 5) includes the formation of solids, so that the
582 modelled water uptake should correspond most closely to the measured "deliquescence scans" of Taylor et
583 al. (2017) (particles exposed to low *RH* then high *RH*).

584 The results for all samples are shown in Figure 8, with the exception of sample S5 (because growth factors
585 of WSOC extract SX5 were not measured). The upper and lower limits of the shaded area, at any given *RH*,
586 represent the two charge balance cases described in section 4.1 (either the cation amounts were adjusted to
587 match the measured total anion charge, or vice versa). The insets show the contributions to the total growth
588 factor of the higher molecular weight WSOC fraction (measured), the WSOC fraction plus the calculated
589 polar organic contribution, and finally all three components. There are several notable features of the plots:

590 First, calculated growth factors are generally lower than measured values. This is especially the case for
591 sample S2, which has the lowest inorganic fraction of all the samples (see Figure 1a,c) and consequently the
592 lowest predicted growth factor. In Figure 9 we show the measured growth factors for all samples at three
593 selected *RH*, plotted against the mol % of inorganic solutes. There is no apparent relationship between the
594 two, which is not what would be expected. We investigated this behaviour further by plotting the same
595 quantities, but using the calculated growth factors, in Figure 9b. To this we added the calculated growth
596 factors, for 80% *RH*, of a mixture of 1 mole of $(\text{NH}_4)_2\text{SO}_4$ and 1 mole of organic material, which is assumed
597 to take up water according to Raoult's law and has a dry density and molar volume (when dissolved) the same
598 as the WSOC material. The two dashed lines on the plot correspond to 40% dissolved organic material, and
599 100%. The results in Figure 9b show, first, that there is essentially no relationship expected between mole %
600 of inorganic solutes and growth factor at 60% *RH*, because a significant fraction is calculated to be solid at
601 this point. This is in agreement with the measured values for the same *RH*, shown in Figure 9a. At 80% *RH*
602 the calculated *GF* for the five sample compositions show a relationship with inorganic content with a slope
603 that corresponds quite closely to the two simplified cases (the dashed lines in the Figure 9b). However, this is
604 not the case for the measured growth factors. The main factors that affect these comparisons are: (i)
605 uncertainties in the TDMA measurements, (ii) uncertainties in the composition measurements and the

606 relative amounts of the three aerosol fractions, and (iii) uncertainties in the modeling estimates. A
607 comparison of parts (a) and (b) of Fig. 9 of suggests that it is the measured growth factors for just two
608 composite samples that are most responsible for the apparent lack of correlation: S2 (lowest relative
609 inorganic content, 27.7%) and S1 (highest inorganic content, 55.7%). The differences between measured
610 and modelled growth factors are greatest for sample S2. Sample S2 contains roughly 3x to 4x the amount of
611 WSOC material that the other samples do, but only a typical amount of inorganic ions. The measured
612 growth factor of the WSOC extract SX2 appears anomalously high, in comparison to the other samples, as
613 does that of the total aerosol. This remains unexplained. It is noticeable (Figure 8a) that the *RH* scan points
614 for S1 (highest inorganic content) are compressed at high *RH*. That is an indication that the TDMA raised
615 the *RH* up to 90% relatively slowly, and it is possible that control settings were refined for the later
616 experiments, which would affect the comparability of the results.

617 The total amounts of WSOC plus individually analysed polar organic compounds in the composite samples
618 are obtained using a Shimadzu analyser, the polar organic compounds by a combination of ion
619 chromatography and GC-MS, and the remaining WSOC material by difference. The analytical uncertainties
620 associated with the measurements are given in the notes to Tables 1, 3, and 4. Systematic biases of similar
621 magnitudes, leading to an underestimate of the total proportions of inorganic ions in the composite samples
622 S1 to S6, might be the cause of the under-predicted growth factors for S3, S4, and S6 (but not S2). However,
623 with only 5 samples, it does not seem helpful to speculate about the relative importance of the several
624 sources of uncertainty. Some of the discrepancies likely reflect imprecision inherent in the growth factor
625 measurements, as well as the various elements of the chemical analysis, and the small number of
626 samples. As apparent from the error bars in Figure 9a, the magnitude of deviation from the expected trend
627 is on the order of the analytic uncertainty associated with TDMA and composition measurements. Other
628 uncertainties arise from modelling estimates, and the sensitivity to contamination of microgram samples.

629 Second, there are some differences between the growth factors measured in the two scans for each sample
630 (particularly S4), but in general they are slight. The calculated growth factors also do not show the
631 deliquescence "steps" that are typical of simple purely inorganic systems, except for sample S1. For this
632 sample the increased growth factor at about 72% *RH* is due mostly to the predicted dissolution of
633 $(\text{NH}_4)_2\text{SO}_4$ and $\text{MgSO}_4 \cdot 6\text{H}_2\text{O}$. In all samples numerous salts are predicted to be formed, and the smoothness
634 of most of the growth factor curves in Figure 8 can be attributed to the formation and dissolution of large
635 numbers of salts as *RH* changes. For example, for sample S3, over the *RH* range 80% to 20%, the following
636 solid salts are present: K_2SO_4 , $\text{MgSO}_4 \cdot 6\text{H}_2\text{O}$, $\text{Na}_2\text{SO}_4 \cdot (\text{NH}_4)_2\text{SO}_4 \cdot 4\text{H}_2\text{O}$, $(\text{NH}_4)_2\text{SO}_4$,
637 $\{2,3\}\text{NH}_4\text{NO}_3 \cdot (\text{NH}_4)_2\text{SO}_4$, NH_4Cl , and $\text{MgSO}_4 \cdot \text{H}_2\text{O}$.

638 Third, it is important to remember that the calculation of the water content of the samples as the sum of the
639 three different components (equation 5) introduces an artifact with regard to the formation of the solid salts.
640 This is illustrated in Figure 10 for the case of a hypothetical aerosol containing 1 mole of a soluble
641 "Raoult's law" organic compound and one mole of $(\text{NH}_4)_{1.5}\text{H}_{0.5}\text{SO}_4$ (letovicite). Where the water content
642 associated with the organic compound and salt are calculated separately, the electrolyte fraction contributes
643 nothing to the total water content below the deliquescence point (68% *RH*), whereas in reality the soluble
644 organic fraction of the aerosol provides some water at all *RH* for the salt to dissolve into. This leads to
645 much higher growth factors at moderate to low *RH*, a smoother growth factor curve, and a reduction (i.e.,
646 lower *RH*) of the deliquescence transitions with respect to $(\text{NH}_4)_2\text{SO}_4$ and $(\text{NH}_4)_{1.5}\text{H}_{0.5}\text{SO}_4$. We would not
647 expect such a large effect for the measured samples, because of the apparently small amounts of water
648 associated with the organic fraction, and it would not explain the difference between measured and
649 modelled growth factors at high *RH* for which there are no solid salts.

650 Perhaps more relevant to the present study is the fact that, at high *RH*, equation (5) does not take into
651 account the influence of the additional amounts of WSOC organic material that can be expected to dissolve
652 into the relatively large amount of water associated with the inorganic fraction of the total aerosol. There are
653 two elements to consider: first, the additional volume of water for the organic to dissolve in; second, the
654 change in the activity coefficient (γ) of the organic going from the solution of WSOC material + water (as
655 measured for the SX series of extracts) to a mixture that also contains the inorganic solutes. The latter is
656 probably of lesser importance: a calculation using the Zdanovskii-Stokes-Robinson expression for solute
657 activity coefficients (equation 9 of Clegg et al., 2003) suggests a change from $\gamma(\text{WSOC})$ equal to 0.8 (on a
658 molality basis) in a water-WSOC solution at 80% *RH*, to 0.66 in a mixture containing 1 mole of dissolved
659 $(\text{NH}_4)_2\text{SO}_4$ and 0.5 moles of dissolved WSOC material. In this calculation the ammonium sulphate is used as a
660 surrogate of the more complex inorganic mixtures occurring in the samples, and the 0.5 moles of WSOC
661 material corresponds to a 1:1 mixture (in terms of moles) in which 50% dissolves in water. (The mole %
662 compositions of the samples can be seen in Figure 1c.)

663 The *E-AIM* model was used to investigate the effect of additional dissolution of WSOC organic material in the
664 total aerosol samples (S1-S6), relative to that estimated to occur for SX1 in Figure 7 under various
665 assumptions, as follows. The amounts of water associated with 1 mole of $(\text{NH}_4)_2\text{SO}_4$ and 1 mole of WSOC
666 material (50% dissolved) at 80% *RH* are 9.56 and 2.0 moles, respectively. If we assume a WSOC "dry"
667 density of 1.3 g cm^{-3} , a molar volume of $386 \text{ cm}^3 \text{ mol}^{-1}$ when dissolved (for an average molar mass of 386 g
668 mol^{-1}) then the growth factor of the mixture at 80% is 1.19. Also needed for this calculation is the density of
669 dry $(\text{NH}_4)_2\text{SO}_4$, which is 1.77 g cm^{-3} , and that of its aqueous solution in equilibrium with 80% *RH* which is
670 1.245 g cm^{-3} . This *GF* corresponds quite closely with those predicted for samples S3-S6, which are the ones
671 that are nearest to a 1:1 mixture of inorganic solutes and WSOC material. If a further 25% of the WSOC
672 material dissolves into the water associated with the salt, we calculate an increase in the growth factor of 0.03
673 (to 1.22). This increase, which would bring the predicted *GF* closer to the measured ones for these total
674 samples is significant and is of about the same magnitude as the uncertainty associated with the charge
675 imbalances of the inorganic ion amounts for samples S3 and S4.

676 Fourth, we also calculated growth factors for cases in which a negative charge balance in the measured
677 inorganic ions was corrected by adding H^+ to create an acidic aerosol. One such result, for sample S3, is
678 shown in Figure 8c. This is a typical: a much higher growth factor is seen at low *RH* due to the greater
679 solubility of the acid salts, and this agrees better with the measurements for this sample. However, this is
680 not the case for all samples (too high a growth factor is predicted for sample S1), and the assumption does
681 not explain the lack of a relationship between the inorganic content of the samples and growth factor shown
682 in Figure 9a. Also, a highly acidic aerosol seems likely to be unrealistic for reasons stated earlier.

683 Finally, we note that the *GF* results for samples SX2 and S2 remain an anomaly. Although the inorganic
684 content of this composite aerosol sample, 561 ng m^{-3} , is similar to that of the other samples, the WSOC
685 concentration exceeds that of the other samples by a factor of 3 or greater. Furthermore, this fraction (SX2)
686 has a much higher measured growth factor than WSOC material from the other samples, despite a
687 composition which is similar. This suggests the possibility of chemical contamination of the WSOC
688 fraction. This remains unexplained, but seems unlikely to have occurred during the determination of the
689 growth factors because of the procedures adopted to avoid it. (In between samples, and especially following
690 any calibration with salt, the atomizer was thoroughly purged with pure water. The size distribution generated
691 by the atomizer when filled with a small volume (20 cm^3) of ultrapure water was also periodically measured
692 to ensure that the characteristic diameter of the generated aerosol was at least a factor of 10 lower than
693 generated when the atomizer was filled with the sample solutions. Furthermore, small samples of ultrapure
694 water were run through the system over the course of a few hours, to check that there was no rightward shift

695 in the measured size distribution that would occur if contamination accumulated over the typical time required
696 for the measurement of one sample.

697

698 7. Discussion and Summary

699 Predicting the hygroscopicity of the soluble organic component of an aerosol requires a knowledge of both
700 the compounds present, and their solubility (either from measurements or predictions) at different *RH*. In
701 this work we have explored elements of both of these requirements, and shown that it is possible to assess
702 hygroscopicity in a semi-quantitative way (the hygroscopicity index *HI*) based on the results of FT-ICR MS
703 analysis coupled with predictions of the functional group compositions of molecules. We have used these
704 results, together with the UNIFAC model and measured growth factors of the organic extracts, to determine
705 that the dissolved fraction of the organic material varies smoothly with *RH* (up to 50% or more dissolved at
706 90% *RH* depending on the modelling assumptions used), see Figure 7. Direct quantification of the degree of
707 dissolution of organic aerosol material would be valuable in future studies. The combination of measured
708 hygroscopicities of the organic fraction of the aerosol (Taylor et al., 2017) with the model-predicted water
709 uptake of the inorganic fraction agrees quite well with the measured growth factors (Figure 8) within the
710 uncertainties of the measurements, although with some differences that are noted later in this section. Our
711 results are likely to be particularly relevant to other locations where biogenic secondary organic aerosol
712 dominates. However, the methods and modelling approaches developed here can in principle be applied to
713 any soluble organic aerosol. The main points of our results, focusing particularly on the WSOC material, are
714 summarised and discussed below.

715 The calculated functional group assignments for the high molecular weight WSOC fraction (Table 4), based
716 only on the numbers of double bond equivalents in the WSOC molecules and their formulae, were poorly
717 constrained. It is possible for alcohol and acid groups to make up between 28% and <3% of the total
718 assigned functional groups, depending on the weighting of the assignments, and still obey the constraints
719 (Table A1). However, the FT-ICR MS results in Table 5 suggest that the final set of UNIFAC group
720 assignments in Table A1 (with high weight given to –OH, –COOH, and alkane groups) is the most realistic.

721 The solubility of organic compounds is expected to decrease with the number of carbon atoms (molecular
722 size), and be increased by the presence of polar organic functional groups such as –OH and –COOH. The
723 hygroscopicity index *HI* developed in section 3.3 is used in this work to estimate a ranking of the WSOC
724 compounds, by order of hygroscopicity, taking into account the relative amounts of each that are present.
725 This index determines the order in which WSOC compounds are allowed to dissolve in the calculations of
726 growth factors for different fractional dissolved amounts.

727 The extent to which the WSOC material dissolves in water at different *RH* isn't known directly. Given that
728 many of the molecules are very large – up to about 40 carbon atoms (Table 4) – complete solubility is not
729 expected even at the highest *RH*. This suggests that the solid/liquid phase partitioning of the WSOC
730 partitioning resembles examples 2 (solid core and aqueous solution) or 4 (solid core and two or more liquid
731 phases) from Figure 6. The predicted growth factors of the WSOC material, discussed in section 5 and
732 shown in Figure 7, compare an assumption of "Raoult's law" solubility with UNIFAC calculations for the
733 three different functional group assignments (maximise the total number of groups, minimise it, and give
734 high weight to –OH and –COOH groups). The results show that:

735 (i) In all cases the measured growth factors are consistent with a solubility of the WSOC compounds that
736 varies smoothly with *RH*.

737 (ii) Some of the WSOC material is predicted to remain undissolved even at the highest modelled *RH* (90%),
738 and this fraction is greatest for the two UNIFAC group assignments that include the highest numbers of

739 polar groups. The second of these, in which high weight is assigned to alkane, –OH, and –COOH functional
740 groups, yields the assignments that are most consistent with FT-ICR MS measurements.

741 (iii) The highest fractional solubilities, and therefore the *lowest* hygroscopicities per mole of dissolved
742 WSOC material, are predicted for the Raoult's law case (Figure 7d) and the one in which the total number
743 of assigned functional groups is minimised (Figure 7a). This is the assignment in which –OH and –COOH
744 groups make up less than 3% of the total (Table A1), and contrasts to the two other cases in which UNIFAC
745 was used and for which the proportions of –OH and –COOH groups are much higher (Figure 7b,c).

746 Our finding of partial solubility has implications for aerosol modelling. Riipinen et al. (2015), in a
747 theoretical study, have demonstrated that the solubility of aerosol organic material is an important factor in
748 controlling its ability to act as CCN, and Carrico et al. (2008) have shown that the hygroscopicity of smoke
749 extracts from biomass burning, determined by HTDMA and by CCN measurements, are closely related (to
750 within $\pm 20\%$ in the κ parameter of Petters and Kreidenweis (2007)). Figure 11 shows the soluble fractions
751 of WSOC material for which the calculated growth factor agrees with the measured values, for two cases:
752 Raoult's law behaviour of the aqueous phase (as in Figure 7d), and for the UNIFAC calculation (from
753 Figure 7c). In both cases the soluble fraction varies approximately linearly with RH (from a soluble fraction
754 of zero at 10% RH), and implies that it remains below unity even close to 100% RH . Converting from a
755 mole to a mass basis, the soluble fraction for the Raoult's law case in Figure 11 is given by $0.671(RH - 0.1)$.
756 These fractions correspond to the "flat" solubility distribution shown in Figure 2 of Riipinen et al. (2015):
757 equal mass fractions of material, each with a defined solubility, covering a logarithmic range of solubilities
758 up to 1000 g dm^{-3} .

759 It might be expected that this simple relationship between soluble fraction and RH applies, with different
760 slopes, to other types of multicomponent organic aerosol material. If so, the only extra information needed
761 in order to estimate a growth factor is the density of the solid (undissolved) organic material, the density of
762 the aqueous solution, and an average molar mass of the organic material. The relationship could also be
763 used to derive a κ parameter, as described for material of limited solubility by Petters and Kreidenweis
764 (2008) (and with an RH -dependent value of function $H(x_i)$ in their equation 6). We suggest this as a subject
765 for further research.

766 The individually analysed organic acids and sugar alcohols (Samburova et al., 2013) constitute a significant
767 fraction of the total aerosol in all samples. The calculated growth factors of these compounds are predicted
768 to be intermediate between those of the WSOC fraction, and typical inorganic salts (Figure 5). At 80% RH
769 the predicted GF in the figure is about 1.19 compared to a measured 1.08 for the WSOC material, but a
770 much higher 1.48 for ammonium sulphate. These acids and sugar alcohols have known functional group
771 compositions and much simpler structures than the WSOC compounds, and the UNIFAC model is better
772 suited to modelling their properties. However, in the calculations of the hygroscopicity of this fraction of
773 the aerosol, in Figures 5 and 8, we have assumed complete dissolution of all compounds at all RH . Thus it
774 is probable that the predicted contributions to the growth factors of the total aerosol, shown in the insets to
775 the plots in Figure 8, are maximum values.

776 The growth factors of the total aerosol material, see section 6, were modelled as the sum of the *measured*
777 water uptake of the WSOC fraction, and the *predicted* water uptake of the individually determined organic
778 compounds (discussed in the previous paragraph) and the inorganic ions. The calculations for the ions
779 assumed the equilibrium formation of solid salts. The measured growth factors, shown in Figure 8, are for
780 both "deliquescence" and "efflorescence" scans and are designed to map the lower and upper legs of any
781 hysteresis loop. They are generally quite similar except for sample S4 above about 90% RH . There is some
782 uncertainty in the calculations of the contribution of the ions to total water uptake because of charge
783 imbalances between the total measured cations and total measured anions, which were compensated based
784 on various different assumptions. With the exception of sample S1, the predicted growth factors at

785 moderate to high RH tend to be somewhat lower than measured, although in reasonable overall agreement.
786 The method of estimating the total water content of the aerosol as the sum of that associated with the three
787 fractions (equation 5) does not explicitly take into account the increased dissolution of soluble solids into
788 the larger volume of water that would be expected. An estimate of the effect of increased WSOC
789 dissolution suggested that the growth factors at 80% RH could be increased by about 0.03. This is
790 significant, but is not sufficient to bring the measured and modelled growth factors in Figure 8 into
791 complete agreement.

792 The results of this study broadly validate the approach taken to modelling the hygroscopicity of a
793 compositionally complex aerosol, containing both inorganic and organic compounds. Our finding that
794 partial solubility of the WSOC material is required to explain measured growth factors, and that this can be
795 represented as a linear function of RH , suggests that quite simple approaches can be used to model its
796 atmospheric effects. Studies to quantify directly the degree of dissolution of the organic fraction of the
797 aerosol at different RH , and the phase(s) in which it is present, would be valuable.

798

799

800

801 **Acknowledgements**

802 The National Science Foundation Division of Atmospheric Sciences collaborative grant AGS-0931431,
803 AGS-0931910, AGS-0931505, and AGS-0931390 supported this work. Any opinions, findings, and
804 conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily
805 reflect the views of the National Science Foundation. The authors thank Peter Atkins and Thomas
806 Kristensen for organization of the field campaign and the collection of samples, and Doug Lowenthal for
807 contributions to earlier parts of the project (sampling and analysis).

808

810 **Estimation of the Functional Group Compositions**

811 We estimated the compositions of the high molecular weight WSOC material in the samples in terms of the
812 UNIFAC functional groups in the following ways. First, it is necessarily assumed that the only functional
813 and structural groups present are those available within UNIFAC. For simplicity, it is also assumed that the
814 molecules consist either of chains of carbon atoms (with branches, if necessary, but not aliphatic rings), or a
815 single aromatic ring with either one or two carbon chains attached. The determination of the possible
816 functional and structural group compositions for each molecule was formulated as a constrained integer
817 minimisation problem in which each molecule is described using the minimum number of UNIFAC groups,
818 subject to the constraints that there must be no unoccupied bonds, and the numbers of atoms and double
819 bond equivalents must be correct for each molecule. The problem was solved using a "branch and bound"
820 linear programming method (routine H02BBF of the Numerical Algorithms Group Fortran Library (NAG,
821 2013)).

822 The available information regarding the molecules is insufficient to provide unique solutions in terms of the
823 assigned functional groups, and minimising the total number of groups tends to favour those that contain
824 large numbers of atoms (such as acetate, for example). We therefore carried out two additional calculations,
825 for the WSOC molecules in samples S1, to explore the variability of estimated composition. In the first
826 calculation, the group assignments were carried out so as to maximise the total number of groups in each
827 molecule, thus favouring the presence of the smaller groups containing fewer atoms. In the second
828 calculation, the group assignments were weighted in order to favour alkane, alcohol, and acid groups and
829 thus both describe the molecules in a functionally simple way and maximise the occurrence of the two polar
830 groups that most strongly promote hygroscopicity. The results are shown in Table A1. In the base case
831 (minimising the number of groups needed to describe each molecule) the majority of the oxygen atoms in
832 the molecules are assigned to acetate and ether groups and most of the rest to ketones. When the number of
833 groups per molecule is maximised (see the second column of results in Table A1) the picture is very
834 different: alcohol and aldehyde groups are now almost 40% of the total, and most of the rest are alkane
835 groups. In the final case, in which high weights are given to alkane, alcohol and acid groups, the two polar
836 groups account for almost 29% of the total number of groups. Another notable feature of the result is that
837 the proportion of assigned alkene groups varies relatively little – from about 7 % to 11 % across the three
838 results. Table A2 shows the group assignments for an arbitrarily chosen molecule ($C_{11}H_{18}O_6$), illustrating –
839 in a single example – how widely they differ between the three cases.

840 Some of the characteristics of the higher molecular weight WSOC material in SX4 are summarised in Table
841 1 of Mazzoleni et al. (2012). The frequency of aromatic molecules in the sample, using the aromaticity
842 index of Koch and Dittmar (2006), is only 45 out of a total of 3737 assigned formulae (1.2%). Using the
843 assumption of a "chain" (non-aromatic) molecule, successful UNIFAC functional group assignments were
844 made in this study in all but 0.81% to 3.5% (average: 1.6%) of molecules in samples SX1 to SX6. For the
845 assumption that each molecule contained an aromatic ring (where the number of DBE allowed it),
846 successful assignments were made in all but 7% to 12.6% (average, 8.8%) of cases. These results are for the
847 base case group assignments. While they are qualitatively consistent with the finding of Mazzoleni et al.
848 noted above, the fact that about 90% of the molecules could be assigned group compositions *including*
849 aromatic rings confirms that the assignments are quite weakly constrained by the available information
850 (numbers of each atom present, and double bond equivalents). In the calculations that follow we have first
851 of all accepted the successful chain-based group assignments. Where these were unsuccessful, assignments
852 that include an aromatic ring were adopted where possible. The total mole fraction of WSOC material that
853 was successfully described in this way, for all six samples, ranged from 0.993 to 0.999 (average: 0.9977).

854 The mean numbers of each type of functional group per molecule, and the deviations of each sample from
855 this mean, are listed in Table A3 for the base case group assignments. The results for the different samples,
856 SX1-SX6, are broadly similar. The presence of large numbers of O-containing groups (acetate, ether,
857 ketone) reflects the high degree of oxygenation of most molecules.

858 The modified aromaticity index (AI_{mod}) of Koch and Dittmar et al. (2006) was used to estimate the extent of
859 carbon-carbon unsaturation for the molecular formulae, because the ultrahigh resolution MS/MS analysis
860 cannot otherwise provide specific information about the alkane, alkene, and aromatic functional groups in the
861 WSOC molecules. The index assumes that 50% of the oxygen in a molecular formula contributes a single
862 "unsaturation" in the form of an aldehyde, ketone, or carboxylic acid. The remaining fraction of saturations
863 are then assumed to be due to carbon-carbon bonds and are classified as aliphatic, olefinic, aromatic, or
864 condensed aromatic structures. Overall, the most common AI_{mod} classifications were aliphatic and olefinic,
865 both of which have alkane groups. Thus, it is logical that alkane and alkene groups would represent a
866 significant component of the WSOC. Our group assignments, described above, and AI_{mod} both predict that
867 aromatic groups are the least prevalent. The UNIFAC predictions for aromatic, alkane and alkene groups seem
868 reasonable for all three parameter sets in Table A1.

869 The ultrahigh resolution MS/MS analysis provides insight regarding the polar functional groups associated
870 with the studied precursor molecular formulas (LeClair et al., 2012). Carboxyl ($-COOH$) and hydroxyl ($-OH$)
871 functional groups are observed as neutral losses of CO_2 and H_2O . Thus, the predicted UNIFAC functional
872 groups can be compared to the 720 studied CHO precursor ions and their product ions. In case 1 (see the first
873 column of results in Table A1), where the total groups are minimized, relatively low percentages of hydroxyl
874 and carboxyl functional groups are predicted, which is not supported by the MS/MS neutral loss analysis.
875 Case 2, where the total number of groups are maximized, shows a significant number of hydroxyl functional
876 groups (occurring in 93% of all molecules), which is consistent with the MS/MS results. However, the
877 numbers of carboxyl groups (0.23% of all molecules) are much lower than inferred from the MS/MS analysis
878 as shown in Table 5. In case 3 the alkane, hydroxyl, and carboxyl groups are given high weight in the group
879 assignment calculation, and consequently are predicted to be present in much greater numbers (see the last
880 column of results in Table A1). This is more consistent with the results of the MS/MS analysis, although the
881 predicted numbers of hydroxyl functional groups are low relative to the observed H_2O losses in MS/MS
882 analysis. This may reflect the fact that the MS/MS fragmentation cannot distinguish the neutral loss of H_2O
883 from an independent hydroxyl group ($R-OH$, where R is not a carbonyl carbon) from that of a carboxylic acid
884 group ($R-C(O)-OH$). This means that some of the H_2O losses are likely due to carboxyl functional groups that
885 easily lose an OH. This is supported by the observation that nearly every precursor (all but 59) that shows an
886 H_2O loss also shows a CO_2 loss. Overall, it seems clear that minimizing the functional groups produces the
887 least realistic results, while maximizing the functional groups and weighting certain functional groups
888 produces results more consistent with the ultrahigh resolution MS/MS analysis.

889

890

891

REFERENCES

- 892
893
894 K. Balslev and J. Abildskov (2002) UNIFAC parameters for four new groups. *Ind. Eng. Chem. Res.* **41**, 2047-
895 205.
- 896
897 O. Boucher, D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H.
898 Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang (2013) Clouds and
899 Aerosols. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth*
900 *Assessment Report of the Intergovernmental Panel on Climate Change*, eds. T. F. Stocker, D. Qin, G.-K.
901 Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge
902 University Press.
- 903
904 C. M. Carrico, M. D. Petters, S. M. Kreidenweis, J. L. Collett Jr., G. Engling, and W. C. Malm (2008) Aerosol
905 hygroscopicity and cloud droplet activation of extracts of filters from biomass burning experiments. *J.*
906 *Geophys. Res.* **113**, D08206, doi:10.1029/2007JD009274.
- 907
908 S. L. Clegg, P. Brimblecombe, and A. S. Wexler (1998) A thermodynamic model of the system $H^+ - NH_4^+ -$
909 $SO_4^{2-} - NO_3^- - H_2O$ at tropospheric temperatures. *J. Phys. Chem. A* **102**, 2137-2154.
- 910
911 S. L. Clegg and J. H. Seinfeld (2006a) Thermodynamic models of aqueous solutions containing inorganic
912 electrolytes and dicarboxylic acids at 298.15 K. II. Systems including dissociation equilibria. *J. Phys. Chem. A*
913 **110**, 5718-5734.
- 914
915 S. L. Clegg and J. H. Seinfeld (2006b) Thermodynamic models of aqueous solutions containing inorganic
916 electrolytes and dicarboxylic acids at 298.15 K. I. The acids as non-dissociating components. *J. Phys. Chem.*
917 *A* **110**, 5692-5717.
- 918
919 S. L. Clegg, J. H. Seinfeld and P. Brimblecombe (2001) Thermodynamic modelling of aqueous aerosols
920 containing electrolytes and dissolved organic compounds. *J. Aerosol. Sci.* **32**, 713-738.
- 921
922 S. L. Clegg, J. H. Seinfeld, and E. O. Edney (2003) Thermodynamic modelling of aqueous aerosols containing
923 electrolytes and dissolved organic compounds. II. An extended Zdanovskii-Stokes-Robinson approach. *J.*
924 *Aerosol. Sci.* **34**, 667-690.
- 925
926 S. L. Clegg and J. M. Simonson (2001) A BET model of the thermodynamics of aqueous multicomponent
927 solutions at extreme concentration. *J. Chem. Thermodyn.* **33**, 1457-1472.
- 928
929 S. L. Clegg and A. S. Wexler (2011) Densities and apparent molar volumes of atmospherically important
930 electrolyte solutions. 1. The solutes H_2SO_4 , HNO_3 , HCl , Na_2SO_4 , $NaNO_3$, $NaCl$, $(NH_4)_2SO_4$, NH_4NO_3 , and
931 NH_4Cl from 0 to 50 °C, including extrapolations to very low temperature and to the pure liquid state, and
932 $NaHSO_4$, $NaOH$, and NH_3 at 25 °C. *J. Phys. Chem. A* **115**, 3393-3460.
- 933
934 A. Fredenslund, R. L. Jones, J. M. Prausnitz, (1975) Group-contribution estimation of activity coefficients in
935 non-ideal liquid mixtures *AIChEJ.*, **21**, 1086-1099.
- 936
937 Freedman, M. A. (2017) Phase separation in organic aerosol. *Chemical Society Reviews*, **46**(24), 7694-7705.
- 938
939 G. S. Girolami (1994) A Simple "Back of the Envelope" Method for Estimating the Densities and
940 Molecular Volumes of Liquids and Solids, *J. Chem. Educ.* **71**, 962-964.
- 941
942 A. G. Hallar, D. H. Lowenthal, S. L. Clegg, V. Samburova, N. Taylor, L. R. Mazzoleni, B. K. Zielinska, T.
943 B. Kristensen, G. Chirokova, I. B. McCubbin, C. Dodson, D. Collins (2013) Chemical and hygroscopic
944 properties of aerosol organics at Storm Peak Laboratory. *J. Geophys. Res.(Atmospheres)* **118**, 1-13,
945 doi:10.1002/jgrd.50373.
- 946
947 M. Hallquist, J. C. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, J. Dommen, N. M.
Donahue, C. George, A. H. Goldstein, J. F. Hamilton, H. Herrmann, T. Hoffmann, Y. Iinuma, M. Jang, M.

- 948 E. Jenkin, J. L. Jimenez, A. Kiendler-Scharr, W. Maenhaut, G. McFiggans, Th. F. Mentel, A. Monod, A. S.
949 H. Prevot, J. H. Seinfeld, J. D. Surratt, R. Szmigielski, and J. Wildt (2009) The formation, properties and
950 impact of secondary organic aerosol: current and emerging issues. *Atmos. Chem. Phys.* **9**, 5155-5236.
- 951 H. K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller, and J Gmehling (1991) Vapor-liquid equilibria by
952 UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **30**, 2352-2355.
953
- 954 Y.-F. Hu (2000) An empirical approach for estimating the density of multicomponent aqueous solutions
955 obeying the linear isopiestic relation, *J. Solut. Chem.* **29**, 1229-1236.
956
- 957 Mark Z. Jacobson (1999) *Fundamentals of Atmospheric Modelling*, Cambridge University Press, 656 pp.
958
- 959 M. Kanakidou, J. H. Seinfeld, S. N. Pandis, I. Barnes, F. J. Dentener, M. C. Facchini, R. Van Dingenen, B.
960 Ervens, A. Nenes, C. J. Nielsen, E. Swietlicki, J. P. Putaud, Y. Balkanski, S. Fuzzi, J. Horth, G. K.
961 Moortgat, R. Winterhalter, C. E. L. Myhre, K. Tsigaridis, E. Vignati, E. G. Stephanou, and J. Wilson (2005)
962 Organic aerosol and global climate modelling: a review, *Atmos. Phys. Chem.* **5**, 1053-1123.
963
- 964 B. P. Koch and T. Dittmar (2006) From mass to structure: an aromaticity index for
965 high-resolution mass data of natural organic matter, *Rapid Communications in Mass Spectrometry* **20**, 926-
966 932. See also the erratum published in 2016 in volume 30, page 250, of the same journal.
967
- 968 T. B. Kristensen, H. Wex, B. Nekat, J. K. Nøjgaard, D. van Pinxteren, D. H. Lowenthal, L. R. Mazzoleni, K.
969 Dieckmann, C. B. Koch, T. F. Mentel, H. Herrmann, A. G. Hallar, F. Stratmann, and M. Bilde (2012)
970 Hygroscopic growth and CCN activity of HULIS from different environments, *J. Geophys. Res.* **117**, D22203,
971 doi:10.1029/2012JD018249.
972
- 973 J. P. LeClair, J. L. Collett, and L. R. Mazzoleni (2012) Fragmentation analysis of water-soluble atmospheric
974 organic matter using ultrahigh-resolution FT-ICR mass spectrometry. *ES&T* **46**, 4312-4322.
975
- 976 Numerical Algorithms Group (2013) *The NAG Fortran Library, Mk. 24*, Oxford, (<http://www.nag.co.uk>).
977
- 978 S. T. Martin (2000) Phase transitions of aqueous atmospheric particles. *Chem. Rev.* **100**, 3403-3453.
979
- 980 L. R. Mazzoleni, P. Saranjampour, M. M. Dalbec, V. Samburova, A. G. Hallar, B. Zielinska, D. H. Lowenthal,
981 and S. D. Kohl (2012) Identification of water-soluble organic carbon in non-urban aerosols using ultrahigh-
982 resolution FT-ICR mass spectrometry: organic anions, *Environ. Chem.* **9**, 285-297.
983
- 984 M. D. Petters and S. M. Kreidenweis (2007) A single parameter representation of hygroscopic growth and
985 cloud condensation nucleus activity, *Atmos. Chem. Phys.*, **7**, 1961-1971.
986
- 987 M. D. Petters and S. M. Kreidenweis (2008) A single parameter representation of hygroscopic growth and
988 cloud condensation nucleus activity – Part 2: Including solubility, *Atmos. Chem. Phys.*, **8**, 6273-6279.
989
- 990 C. A. Pope III, and D. W. Dockery (2006) Health effects of fine particulate air pollution: lines that connect,
991 *J. Air Waste Manage.*, **56**, 709-742.
992
- 993 I. Riipinen, N. Rastak, and S. N. Pandis (2015) Connecting the solubility and CCN activation of complex
994 organic aerosols: a theoretical study using solubility distributions, *Atmos. Chem. Phys.* **15**, 6305-6322.
995
- 996 V. Samburova, A. G. Hallar, L. R. Mazzoleni, P. Saranjampour, D. H. Lowenthal, S. D. Kohl, and B.
997 Zielinska (2013) Composition of water-soluble organic carbon in non-urban atmospheric aerosol collected at
998 the Storm Peak Laboratory, *Environ. Chem.* **10**, 370-380.
999
- 1000 M. Semmler, B. P. Luo, T. Koop (2006) Densities of liquid $\text{H}^+ / \text{NH}_4^+ / \text{SO}_4^{2-} / \text{NO}_3^- / \text{H}_2\text{O}$ solutions at
1001 tropospheric temperatures. *Atmos. Environ.*, **40**, 467-483.
1002

- 1003 J. H. Seinfeld and S. N. Pandis (2006) *Atmospheric Chemistry and Physics: From Air Pollution to Climate*
1004 *Change*, 2nd Edn. (revised), Wiley-Blackwell, 1232 pp.
1005
- 1006 M. Shiraiwa, Y. Li, A. P. Tsimpidi, V. A. Karydis, T. Berkemeier, S. N. Pandis, J. Lelieveld, T. Koop, and
1007 U. Poschl (2017) Global distribution of particle phase state in atmospheric secondary organic aerosols.
1008 *Nature Communications*, **8**, 15002 (doi: 10.1038/ncomms15002).
- 1009 J. D. Smith, V. Sio, L. Yu, Q. Zhang, and C. Anastasio (2014) Secondary organic aerosol production from
1010 aqueous reactions of atmospheric phenols with an organic triplet excited state. *ES&T* **48**, 1049-1057.
- 1011 R. H. Stokes, and R. A. Robinson (1966) Interactions in aqueous nonelectrolyte solutions: I. Solute-solvent
1012 equilibria. *J. Phys. Chem.* **70**, 2126-2130.
1013
- 1014 S. Suda Petters, D. Pagonis, M. S. Claflin, E. J. T. Levin, M. D. Petters, P. J. Ziemann, and S. M.
1015 Kreidenweis (2017) Hygroscopicity of organic compounds as a function of carbon chain length and
1016 carboxyl, hydroperoxy, and carbonyl functional groups. *J. Phys. Chem. A* **121**, 5164-5174.
1017
- 1018 S. Suda Petters, S. M. Kreidenweis, A. P. Grieshop, P. J. Ziemann, and M. D. Petters (2019) Temperature-
1019 and humidity-dependent phase states of secondary organic aerosols. *Geophys. Res. Lett.*,
1020 doi:10.1029/2018GL080563
1021
- 1022 N. F. Taylor, D. R. Collins, D. H. Lowenthal, I. B. McCubbin, A. G. Hallar, V. Samburova, B. Zielinska, N.
1023 Kumar, and L. R. Mazzoleni (2017) Hygroscopic growth of water soluble organic carbon isolated from
1024 atmospheric aerosol collected at US national parks and Storm Peak Laboratory. *Atmos. Chem. Phys.* **17**,
1025 2555-2571.
1026
- 1027 C. H. Tong, S. L. Clegg and J. H. Seinfeld (2008) Comparison of activity coefficient models for
1028 atmospheric aerosols containing mixtures of electrolytes, organics, and water. *Atmospheric Environment* **42**,
1029 5459-5482.
1030
- 1031 T. V. Vu, J. M. Delgado-Saborit, and R. M. Harrison (2015) A review of hygroscopic growth factors of
1032 submicron aerosols from different sources and its implication for calculation of lung deposition efficiency of
1033 ambient aerosols. *Air Qual. Atmos. Health* **8**, 429-440.
1034
- 1035 J. Wang, M. J. Cubison, A. C. Aiken, J. L. Jimenez, and D. R. Collins (2010) The importance of aerosol
1036 mixing state and size-resolved composition on CCN concentration and the variation of the importance with
1037 atmospheric aging of aerosols. *Atmos. Chem. Phys.* **10**, 7267-7283.
1038
- 1039 A. S. Wexler and S. L. Clegg (2002) Atmospheric aerosol models for systems including the ions H^+ , NH_4^+ ,
1040 Na^+ , SO_4^{2-} , NO_3^- , Cl^- , Br^- , and H_2O . *J. Geophys. Res.-Atmos.* **107**, D14, Art. No. 4207, (see
1041 <http://www.aim.env.uea.ac.uk/aim/aim.php>).
1042
- 1043 R. Wittig, J. Lohmann, and J. Gmehling (2003) Vapor-Liquid Equilibria by UNIFAC Group Contribution. 6.
1044 Revision and Extension. *Ind. Eng. Chem. Res.* **42**, 183-188.
1045
- 1046 M. Zark, J. Christophers, and T. Dittmar (2017) Molecular properties of deep-sea dissolved organic matter
1047 are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **191**, 9-15.
1048
- 1049 R. A. Zaveri, R. C. Easter, J. D. Fast, and L. K. Peters (2008) Model for Simulating Aerosol Interactions
1050 and Chemistry (MOSAIC), *J. Geophys. Res.* **113**, D13204, doi:10.1029/2007JD008782.
1051
- 1052 A. Zuend, C. Marcolli, B. P. Luo, and T. Peter (2008) A thermodynamic model of mixed organic-inorganic
1053 aerosols to predict activity coefficients. *Atmos. Chem. Phys.* **8**, 4559-4593.
1054

Table 1. The measured inorganic composition of the total aerosol material in samples S1-S6.

	Unit	S1	S2	S3	S4	S5	S6
Total ions	ng m ⁻³	797.3	561.3	730.4	368.0	181.8	395.8
(as above)	nmol m ⁻³	14.07	10.52	12.03	9.27	3.66	7.10
Cl⁻	nmol m ⁻³	0.125	0.121	0.0719	0.157	0.0674	0.0124
NO₃⁻	nmol m ⁻³	1.340	1.345	2.213	2.032	0.436	0.808
SO₄²⁻	nmol m ⁻³	5.662	3.435	4.709	0.805	1.008	2.518
NH₄⁺	nmol m ⁻³	4.690	3.210	2.279	3.643	1.164	1.838
Na⁺	nmol m ⁻³	0.281	0.381	0.478	0.389	0.152	0.277
K⁺	nmol m ⁻³	1.063	1.167	1.374	1.132	0.362	0.742
Ca²⁺	nmol m ⁻³	0.707	0.665	0.661	0.829	0.348	0.812
Mg²⁺	nmol m ⁻³	0.203	0.191	0.245	0.287	0.124	0.0922
Charge balance^a	%	-47.8	-25.2	-65.3	64.3	4.02	-22.6
^b	%	4.85	13.8	-8.1	84.7	26.9	16.6
^c	%	-14	-1.3	-26.3	74.1	14.7	0.85

Note: the concentrations are given as amounts per m³ of atmosphere sampled. The analytical uncertainties for inorganic ions were: 3-10 % (Cl⁻), 1-23 % (NO₃⁻), 2-16 % (SO₄²⁻), 3-12 % (NH₄⁺), 0.2-1 % (Na⁺), 1-4 % (K⁺), 1-5 % (Ca²⁺), 4-14 % (Mg²⁺).

^a Charge balance is calculated as $(\sum_i nC_i zC_i - \sum_i nA_i |zA_i|)$ where n and z are the number of moles and the charge, respectively, of each cation C and anion A . To calculate the percentages listed in the table, the amounts given by the expression above are divided by the quantity $0.5(\sum_i nC_i zC_i + \sum_i nA_i |zA_i|)$.

^b The charge balance is calculated on the assumption that SO₄²⁻ is partially neutralised, and present in the aerosol as HSO₄⁻.

^c The charge balance is calculated on the assumption that SO₄²⁻ is partially neutralised, and present in the aerosol as H_{0.5}SO₄^{1.5-}.

Table 2. The measured inorganic composition of the aerosol material in extracts SX1-SX6.

		SX1	SX2	SX3	SX4	SX5	SX6
Total ions	ng m ⁻³	10.15	12.02	13.55	8.84	15.00	16.86
(as above)	nmol m ⁻³	0.259	0.248	0.335	0.203	0.307	0.438
Cl⁻	nmol m ⁻³	0	0	0	0	0.124	0.0451
NO₃⁻	nmol m ⁻³	0.0247	0.0432	0.0384	0.0376	0	0.0450
SO₄²⁻	nmol m ⁻³	0.0194	0.0462	0.0301	0.0187	0.0886	0.0251
NH₄⁺	nmol m ⁻³	0.0327	0.0371	0.0388	0.0305	0.0477	0.0471
Na⁺	nmol m ⁻³	0.0526	0.0270	0.0733	0.0204	0.0368	0.0890
K⁺	nmol m ⁻³	0.0566	0.0442	0.0732	0.0367	0.0100	0.115
Ca²⁺	nmol m ⁻³	0.0609	0.0421	0.0671	0.0514	0	0.0582
Mg²⁺	nmol m ⁻³	0.0124	0.0079	0.0139	0.0079	0	0.0139
charge balance^a	%	127.9	42.2	111.6	93.3	-104.5	95.2

Note: the concentrations are given as amounts per m³ of atmosphere sampled.

^a See note (a) in Table 1.

Table 3. Summary of the measured polar organic composition of the aerosol material (all samples).

	Unit	S1	S2	S3	S4	S5	S6
All compounds	ng m ⁻³	312.0	239.4	212.1	213.6	94.3	122.0
(as above)	ng C m ⁻³	122.2	93.5	79.5	81.3	33.5	42.7
(as above)	nmol m ⁻³	2.29	1.85	1.96	1.98	0.909	1.20
6 low MW acids	mass%	29.16	34.71	50.30	47.84	59.38	67.10
24 high MW acids	mass%	11.40	14.29	18.26	18.32	15.52	19.54
7 sugar alcohols	mass%	6.39	7.32	6.59	6.45	10.81	5.23
11 sugars	mass%	53.05	43.68	24.84	27.39	14.29	8.12
		SX1	SX2	SX3	SX4	SX5	SX6
All compounds	ng m ⁻³	32.7	31.6	47.9	47.0	3.0	26.1
(as above)	ng C m ⁻³	14.7	13.4	21.2	19.6	1.38	10.6
(as above)	nmol m ⁻³	0.316	0.307	0.470	0.500	0.023	0.261
6 low MW acids	mass%	38.04	39.87	38.59	48.54	12.00	47.02
24 high MW acids	mass%	53.44	40.75	53.85	41.74	32.65	44.51
7 sugar alcohols	mass%	0.43	2.34	0.83	1.44	8.34	0.0
11 sugars	mass%	8.10	17.04	6.72	8.27	47.01	8.47

Note: the concentrations are given as amounts per m³ of atmosphere sampled. The analytical uncertainties for organic species were: 3-20 % (6 low MW acids), 5-29 % (24 high MW acids), 5-32 % (7 sugar alcohols), 5-19 % (11 sugars).

Table 4. Summary of the organic composition of the WSOC aerosol material determined by FT-ICR MS (all samples).

	Unit	S1	S2	S3	S4	S5	S6
Total WSOC ^a	ng C m ⁻³	524.5	1932.3	600.9	595.7	303.0	489.9
WSOC ^b	ng C m ⁻³	402.3	1838.7	521.4	514.4	269.5	447.2
WSOC ^c	nmol m ⁻³	1.92	8.87	2.68	2.44	1.32	2.26
		SX1	SX2	SX3	SX4	SX5	SX6
Total WSOC ^d	ng C m ⁻³	289.6	1682.6	484.5	547.0	282.2	256.6
WSOC ^e	ng C m ⁻³	275.0	1669.1	463.2	527.4	280.77	246.1
WSOC ^c	nmol m ⁻³	1.32	8.05	2.38	2.51	1.37	1.24
Properties							
No. of structures or molecules		3056	3349	2797	3881	3384	2490
Max. molar mass	g	736.9	719.8	730.8	772.9	758.9	700.8
Min. molar mass	g	116.1	106.1	114.1	102.1	102.1	114.1
Mean molar mass	g	386.1	392.1	361.9	388.3	378.1	368.5
Mean molar volume	cm ³ mol ⁻¹	383.1	381.2	359.9	386.5	375.4	367.0
Max. no. of C		45	37	37	42	45	39
Min. no. of C		3	3	3	3	3	4
Mean no. of C		17.4	17.3	16.2	17.5	17.1	16.5
Fraction in common		0.7570	0.7364	0.8370	0.7688	0.7882	0.8523

Notes: The analytical uncertainty for total WSOC material was ~ 10%.

^a Total water soluble organic carbon in the aerosol determined using the Shimadzu total organic carbon analyser (see section 2).

^b Concentrations of the WSOC material analysed by FT-ICR MS (Mazzoleni et al. 2012), obtained by subtracting from the total (the line above) the concentrations of the individual polar organic molecules determined by IC and GC MS (Samburova et al., 2013).

^c Same as (b) but in molar units.

^d Total water-soluble organic carbon in the resin-extracted composite samples aerosol determined using the Shimadzu total organic carbon analyser

^e Concentrations of the WSOC material analysed by FT-ICR MS, obtained by subtracting from the total (the line above) the concentrations of the individual polar organic molecules determined in the resin-extracted samples by IC and GC MS.

Table 5. Summary of the numbers of precursor formulas showing certain types of losses.

Molecular Group	CO₂ Loss (-COOH)	H₂O Loss (-OH)	Aldehyde Loss	Methoxy Loss	N, S Loss
All	1294 (86.6%)	1360 (91.0%)	940 (48.3%)	972 (65.0%)	689 (46.1%)
CHO	663 (92.1%)	678 (94.2%)	563 (78.2%)	578 (80.3%)	NA
CHNO	369 (95.8%)	367 (95.3%)	205 (53.2%)	218 (56.6%)	378 (98.2%)
CHOS	243 (70.4%)	293 (84.9%)	171 (49.6%)	173 (50.1%)	274 (79.9%)
CHNOS	19 (42.2%)	22 (48.9%)	1 (2.2%)	3 (6.7%)	37 (82.2%)

Notes: the first value is the number of precursor formulas in each category, and the numbers in parentheses are their percentages of the total formulas in each molecular group. The molecular group (leftmost column) is defined as the set of WSOC molecules or compounds containing only the elements listed.

Table A1. Summary of the assignment of UNIFAC functional groups to WSOC molecules in sample SX1, shown as percentages of the total number of groups.

Functional Group ^a	Percentage Contributions of Each Type of Functional Group		
	Minimise total number of groups ^b	Maximise total number of groups ^b	Increased weight given to alkane, alcohol, and acid groups ^b
aromatic	0.08 (0.16)	0.04 (0.16)	0.06 (0.16)
alkane	36.7 (74.9)	48.7 (99.1)	57.9 (96.8)
alkene	11.2 (55.3)	8.31 (63.4)	7.37 (42.2)
alcohol	1.54 (8.10)	24.9 (93.3)	8.38 (42.5)
ketone	5.12 (28.81)	0	1.45 (25.7)
aldehyde	0.18 (1.15)	14.5 (67.3)	0.14 (1.25)
acid	1.20 (10.9)	0.01 (0.23)	20.3 (98.4)
oxide	2.22 (16.6)	0	0
nitro	0.07 (0.72)	0.02 (0.26)	0.07 (0.99)
amide	4.63 (45.1)	2.63 (45.2)	3.33 (45.0)
thiol	0.07 (0.92)	0.45 (10.2)	0.58 (10.2)
sulphide	0.38 (4.84)	0.38 (7.44)	0.47 (7.4)
sulphone	0.93 (11.9)	0	0
acetate	20.9 (87.3)	0	0
ether	14.8 (71.1)	0	0

Notes: the columns list the numbers of groups of each indicated type (for all 3056 molecules in the sample), expressed as a percentage of the total number of assigned groups of all types. All values greater than 5% are in boldface type. The values in parentheses are the numbers of molecules containing one or more of the indicated functional groups (also expressed as percentages).

^a These types correspond to the following "main groups" listed in the supplementary information to Hansen et al. (1991) except as indicated: aromatic - 3 (ACH); alkane - 1 (CH₂); alkene - 2 (C=C); alcohol - 5 (OH); ketone - 9 (CH₂CO); aldehyde - 10 (CH); acid - 20 (COOH); oxide - 53 (see Balslev and Abildskov, 2002); nitro - 26 (CNO₂); amide - 46 (CON); thiol - 29 (CH₃SH); sulphide - 48 (CH₂S); sulphone - 55 (see Wittig et al., 2003); acetate - 11 (CCOO); ether - 13 (CH₂O). Each of the named groups may involve more than one structure, for example alkane means C (bonded to four other atoms), CH (bonded to three other atoms), -CH₂-, and -CH₃.

^b First column of results – the assignment was carried out so as to describe the molecules using the minimum number of groups, with all groups given equal weight; second column – similar to the first column, but the total number of groups was maximised; third column – the total number of groups was minimised, but a high weight was given to alkane, alcohol, and acid functional groups.

Table A2. Examples of group assignments to molecule $C_{11}H_{18}O_6$ (with 3 DBE)

Functional Group	Numbers of Each Type of Functional Group		
	Minimise total number of groups	Maximise total number of groups	Increased weight given to alkane, alcohol, and acid groups
alkane	5	8	8
aldehyde	-	3	-
alcohol	-	3	-
acid	-	-	3
acetate	2	-	-

Notes: the columns list the number of groups of each indicated type that were assigned to the molecule.

Table A3. Average assigned group compositions, per molecule, for the total SX1-6 WSOC material.

UNIFAC Group	Average per molecule	$\Delta S1$	$\Delta S2$	$\Delta S3$	$\Delta S4$	$\Delta S5$	$\Delta S6$
alkane	4.137982	0.5765	0.3748	-0.4686	-0.1224	-0.1489	-0.2114
acetate	2.585348	0.1058	0.0868	-0.0594	-0.0246	-0.0788	-0.0297
ether	1.814833	0.0828	0.2086	-0.1271	-0.1051	-0.0918	0.0326
alkene	1.567942	-0.1320	-0.2113	-0.0165	0.2740	0.1651	-0.0793
ketone	0.589472	0.0681	0.0278	-0.0427	0.0408	-0.0386	-0.0554
amide	0.532353	0.0630	0.2041	-0.1039	0.0364	-0.0580	-0.1415
oxide	0.344804	-0.0597	-0.1025	0.0487	0.0122	0.0318	0.0696
acid	0.225060	-0.0710	-0.0404	0.0400	0.0535	0.0668	-0.0489
alcohol	0.186433	0.0111	0.0630	-0.0087	-0.0392	-0.0009	-0.0253
sulphone	0.141088	-0.0222	-0.0384	0.0193	0.0035	0.0104	0.0273
sulphide	0.067178	-0.0188	-0.0294	0.0102	0.0073	0.0130	0.0176
aldehyde	0.055608	-0.0322	-0.0289	0.0174	0.0018	0.0537	-0.0118
aromatic	0.031884	-0.0220	-0.0157	0.0048	0.0164	0.0191	-0.0027
thiol	0.013206	-0.0040	-0.0048	-0.0053	0.0052	0.0062	0.0026
nitro	0.006489	0.0034	0.0043	-0.0033	0.0005	-0.0013	-0.0036

Notes: these results are for the "base case" group assignments, in which molecules were described using the minimum number of functional groups, without weighting. The second column shows the average group composition per molecule across all samples, and the six columns $\Delta S1$, $\Delta S2$, etc. indicate the deviations of each of the six samples from that mean.

FIGURES

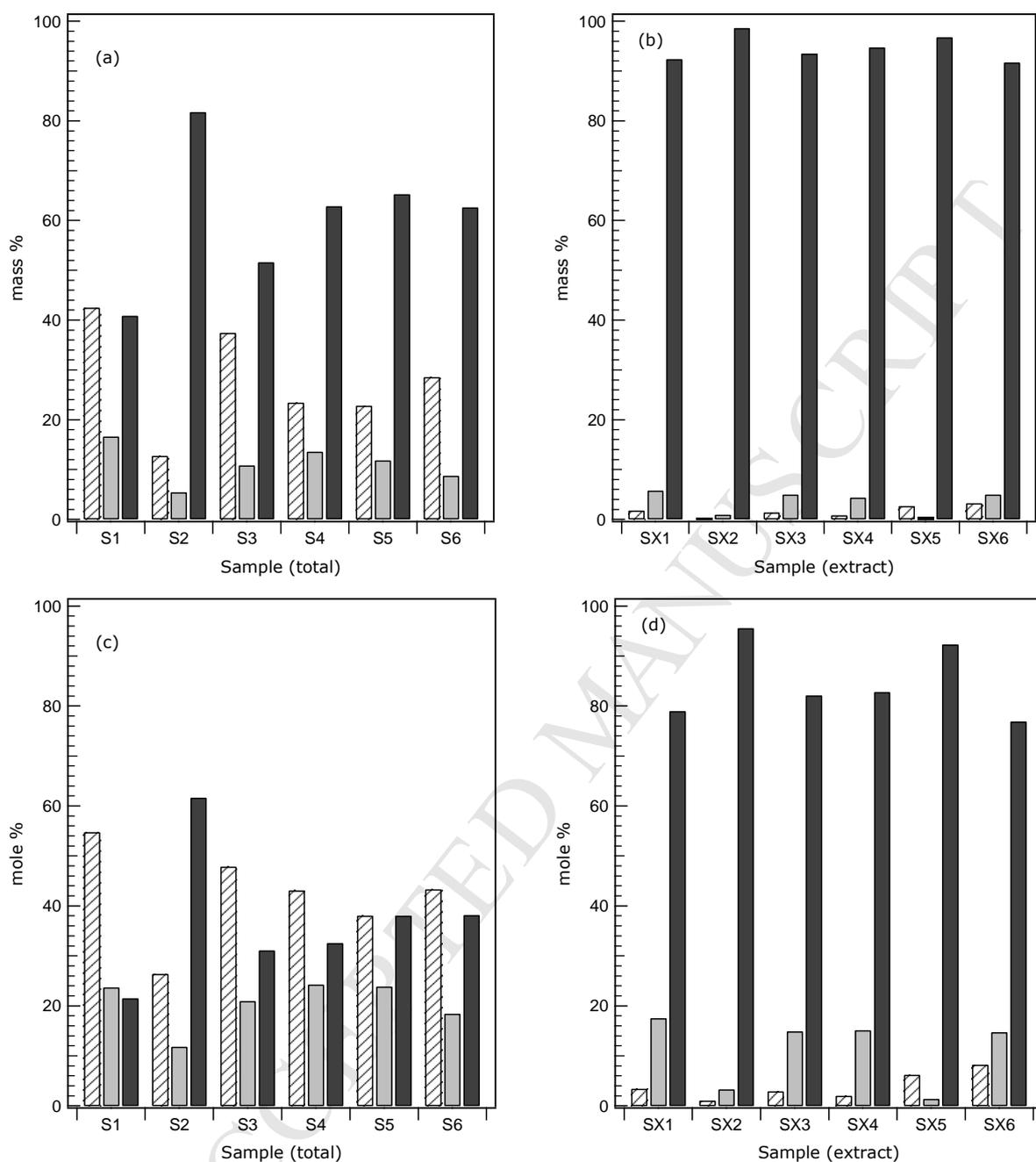


Figure 1. Relative compositions of the total water soluble aerosol material (samples S1-6), and water soluble organic matter extracts (SX1-6), in both mass % (a and b) and mole % (c and d). Key: diagonal lines – inorganic ions; solid grey – individual polar organic compounds; black – high molecular weight water soluble organic carbon (WSOC). The data in this plot are from Tables 3 and 4.

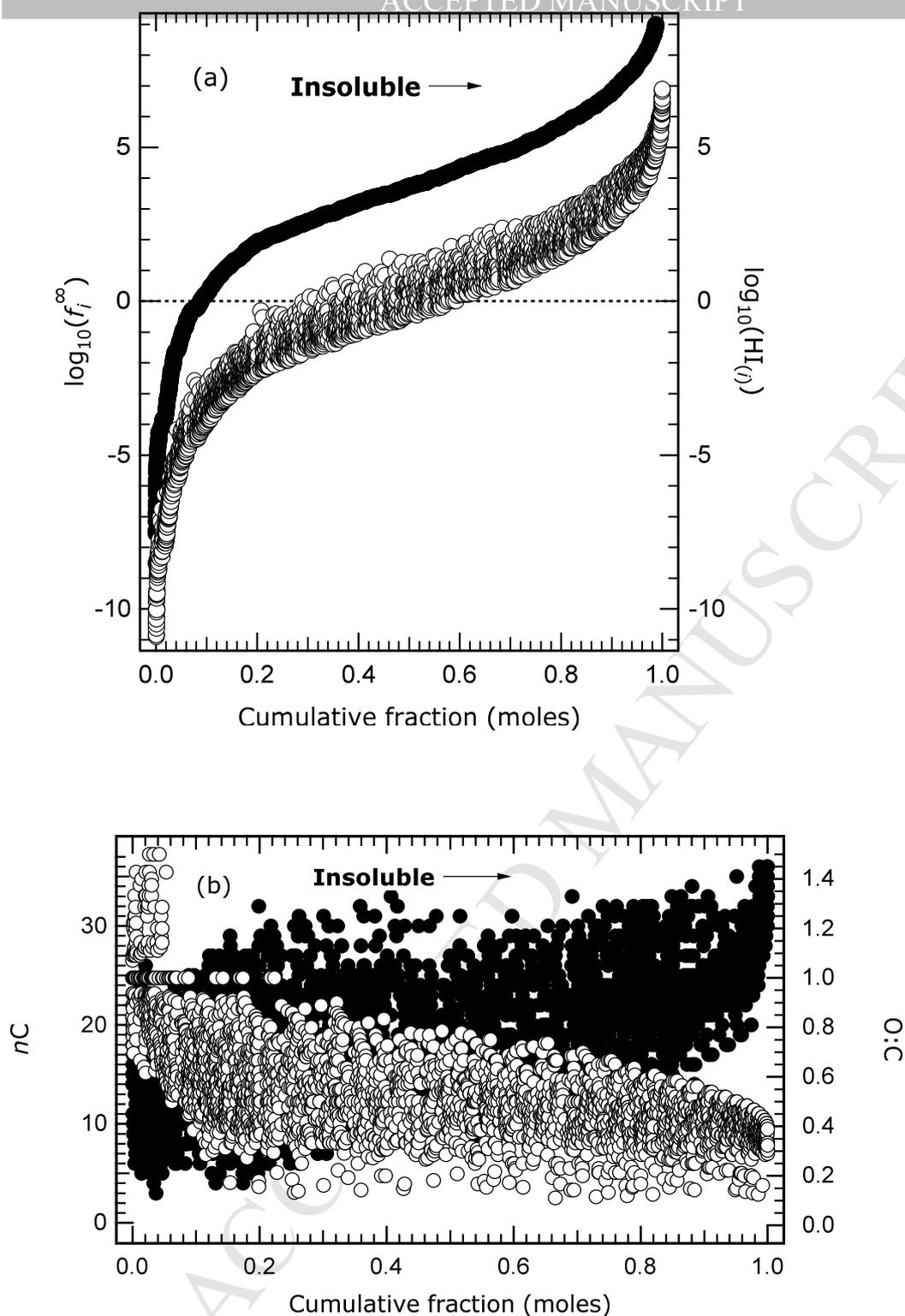


Figure 2. Calculated hygroscopicity index properties of WSOC sample SX1, plotted against the cumulative mole fraction of the sample, ranked in order of increasing values of the index HI . The most hygroscopic compounds are at the left of each plot, and the least hygroscopic compounds at the right. (a) Left axis and solid dots: the logarithm of the calculated activity coefficient of each compound at infinite dilution in water at 25 °C. Right axis and open circles: the index value $HI_{(i)}$ of each compound i . (b) Left axis and solid dots: the number of carbon atoms in each molecule. Right axis and open circles: the ratio of oxygen to carbon atoms in each molecule.

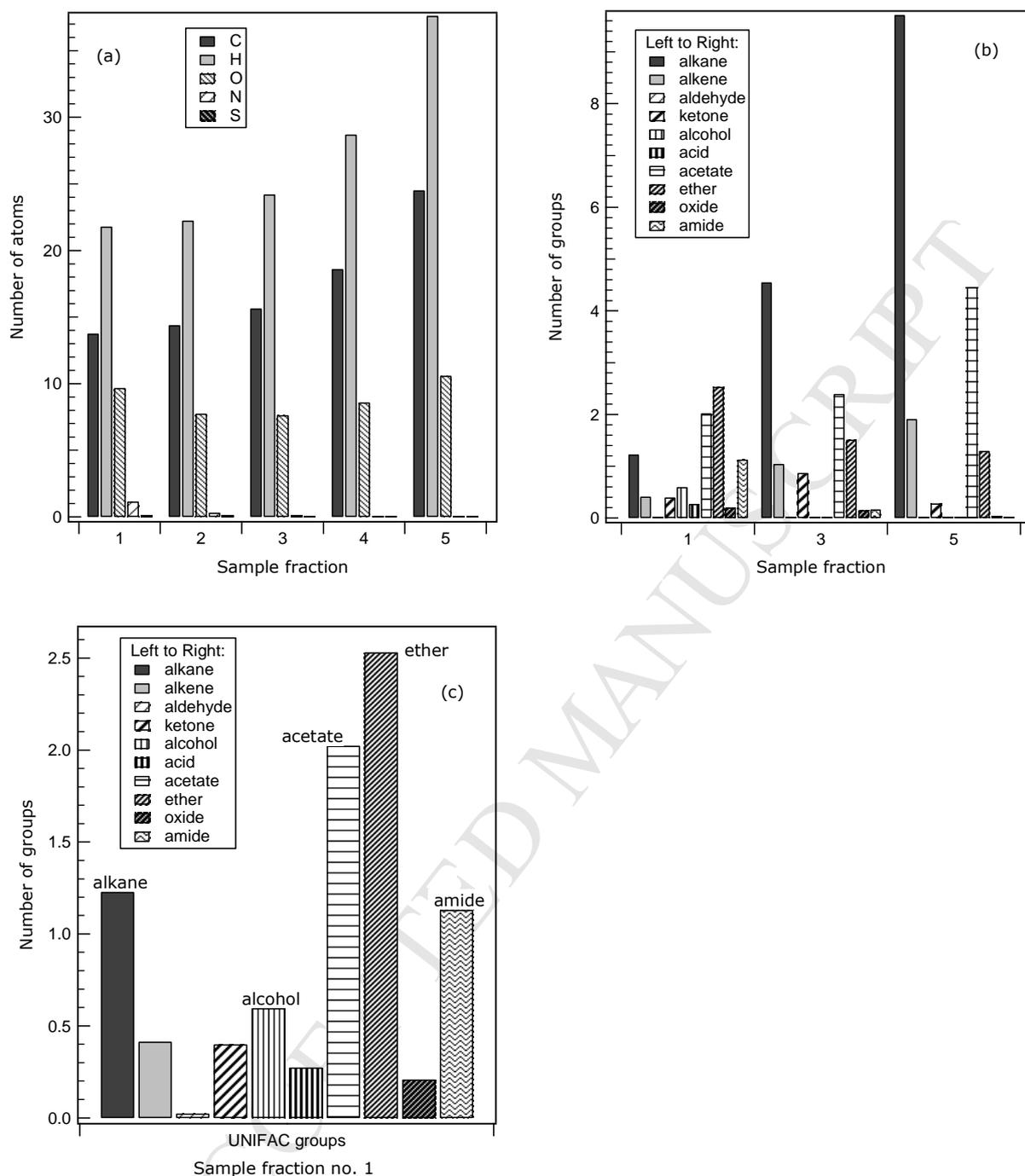


Figure 3. Characteristics of the WSOC fraction in sample SX1, with molecules ranked according to hydroscopicity index and then divided into five groups containing equal numbers of moles of each compound. This is the "base case" in which the functional group assignments are carried out so as to minimise the total number of groups per molecule. Fraction 1 is predicted to be the most soluble, and fraction 5 the least soluble. The ranges of values of $\log_{10}(HI)$ for each fraction are as follows: 1 – less than -1.75; 2 – from -1.75 to 0.0; 3 – from 0.0 to 1.0; 4 – from 1.0 to 2.25; 5 – greater than 2.25. (a) Average formulae of each fraction. (b) Average UNIFAC group composition of fractions 1, 3, and 5. (c) Average UNIFAC group composition of fraction 1 (the same data as shown in b).

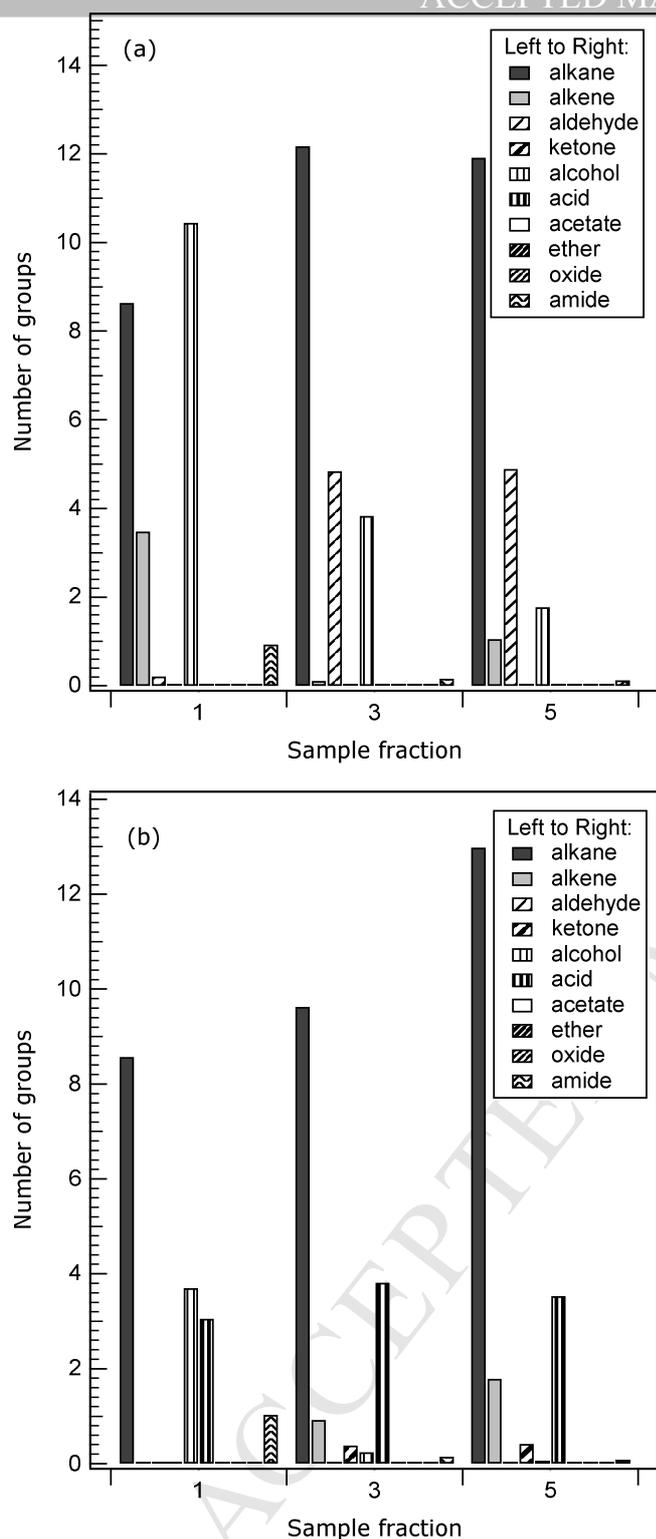


Figure 4. Predicted average UNIFAC group functional composition of the WSOC fraction in sample SX1, with molecules ranked according to hygroscopicity index HI and then divided into five groups containing equal numbers of moles of each compound. Fraction 1 is predicted to be the most soluble, and fraction 5 the least soluble. (a) For the case in which the number of functional groups per molecule was maximised. (b) For the case in which the number of functional groups per molecule was minimised, but high weight was given to alkane, alcohol, and acid groups.

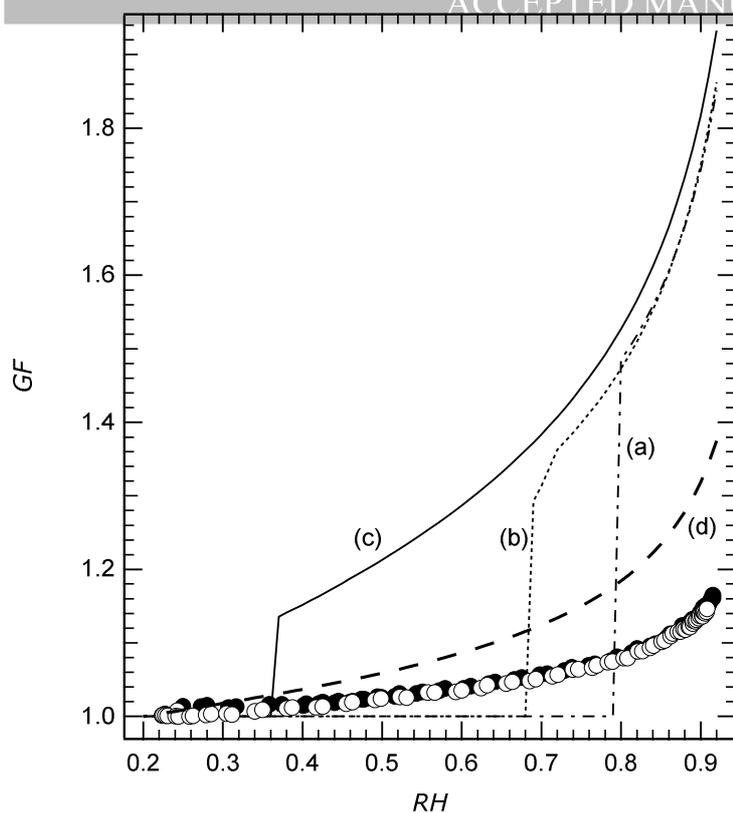


Figure 5. Calculated growth factors (GF) of three ammonium sulphate salts, and polar organic compounds, and measured growth factors of WSOC material SX1 plotted against relative humidity (RH). Symbols (WSOC material): dots – deliquescence (low RH to high RH); open circles – efflorescence (high RH to low RH). Lines: (a) $(\text{NH}_4)_2\text{SO}_4$; (b) $(\text{NH}_4)_3\text{H}(\text{SO}_4)_2$; (c) NH_4HSO_4 ; (d) polar organic compounds assuming a fully liquid mixture at all RH and referenced to the predicted liquid volume at 0.2 RH . The values for the salts were calculated using $E\text{-AIM}$ Model II of Clegg and Wexler (1998) for a temperature of 25 °C, and for the polar organic compounds using UNIFAC, liquid molar volumes estimated using Girolami (1994), and equation (7) of this work.

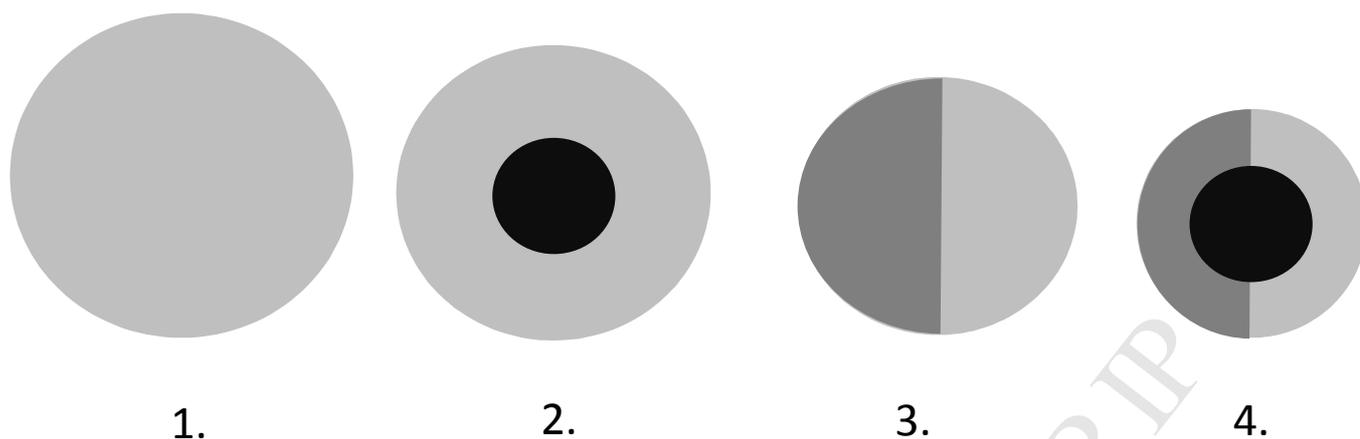


Figure 6. Schematic diagram of possible phases present in WSOC material that has taken up water. (1) A single liquid phase in which all WSOC molecules are miscible with water. (2) A solid core of insoluble material (which may vary in size with the ambient humidity), surrounded by a single liquid phase containing water and that fraction of the WSOC material that is soluble at that relative humidity. (3) Two liquid phases, one containing those molecules that are most soluble in water and a second phase containing largely hydrophobic molecules that are miscible with one another but take up very little water. (4) A combination of cases 2 and 3: an insoluble core, surrounded by hydrophobic and liquid phases in equilibrium with one another.

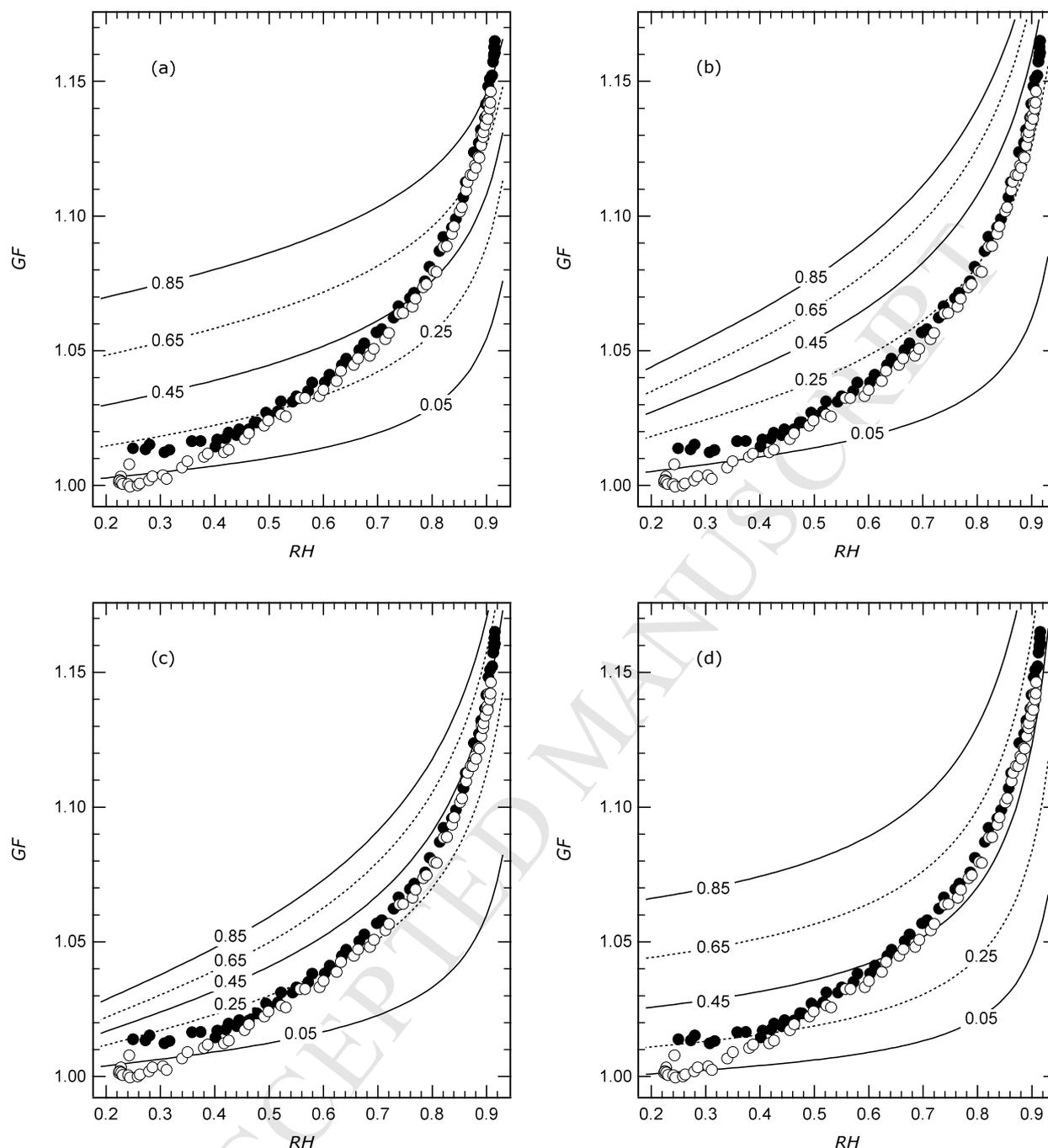


Figure 7. Growth factors (GF) of WSOC material in sample SX1, plotted against ambient relative humidity (RH), expressed as a fraction. Symbols: dots – deliquescence (low RH to high RH); open circles – efflorescence (high RH to low RH); lines - calculated values assuming different fixed mole fractions of aerosol material are soluble (ranging from 0.05 to 0.85, marked on the plots). (a) Calculated GF are for UNIFAC group assignments that minimise the number of functional groups per molecule (the base case); (b) for UNIFAC group assignments that maximise the number of functional groups per molecule; (c) for UNIFAC group assignments that give high weight to alkane, $-OH$, and $-COOH$ functional groups; (d) assuming Raoult's law behaviour of all WSOC molecules. The calculated GF are relative to completely dry material at a relative humidity of 20%, and include the influence of the trace amounts of ions and polar molecules shown in Figure 1(b,d).

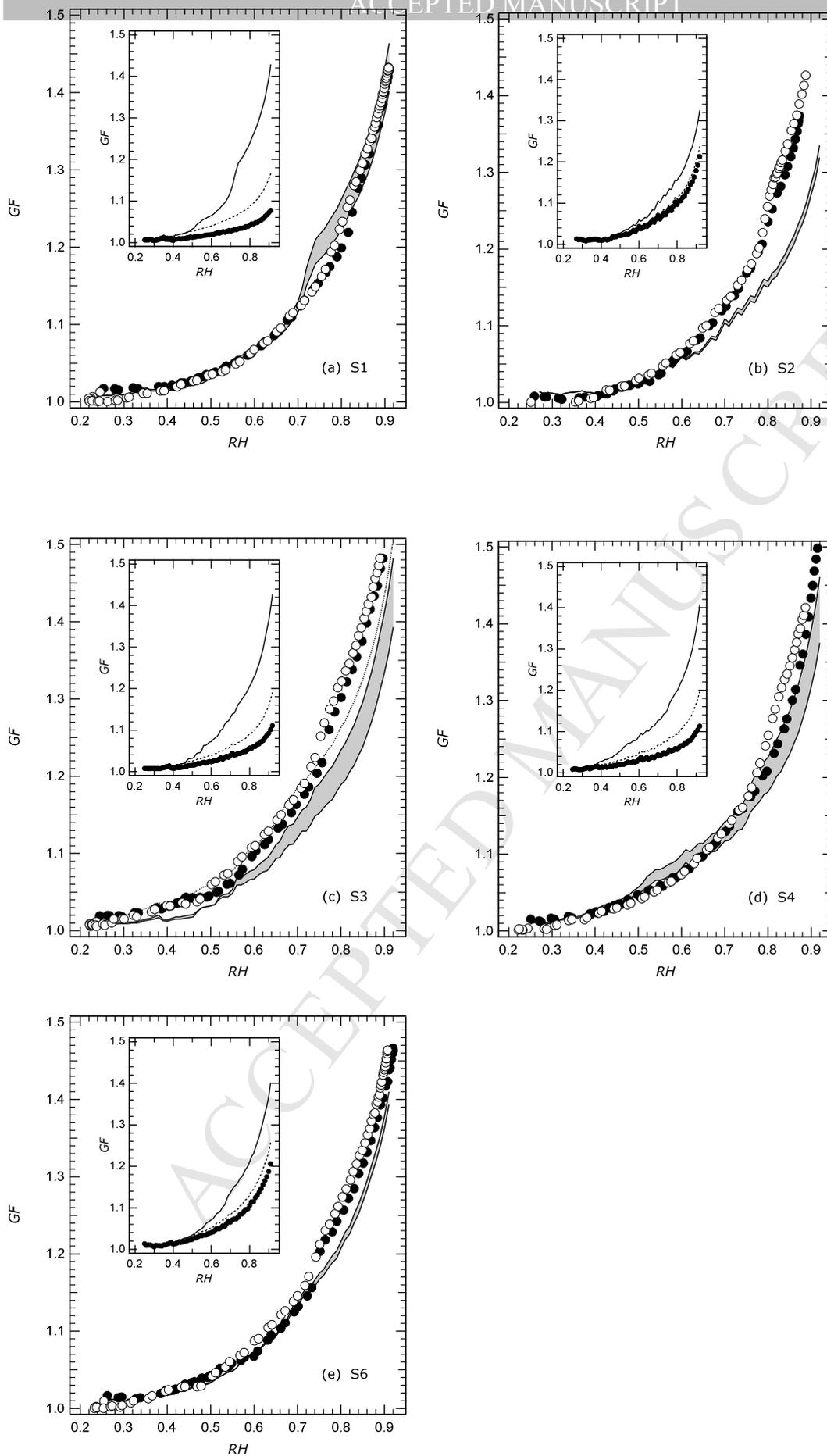


Figure 8. Measured and predicted growth factors (GF) as a function of relative humidity (RH). Symbols: dots – deliquescence measurement (scan from low RH to high RH); open circles – efflorescence

measurement (scan from high *RH* to low *RH*); lines – growth factors calculated as described in the text. The upper and lower solid plotted lines are for the cases in which, (i) the ion amounts have been charge balanced to match the higher of the two totals (i.e. the total ion charge for the cations, or the total for the anions); (ii) they have been charge balanced to match the lower of the two totals. The latter results in a smaller total amount of ionic solutes and hence a lower *GF*. The fine dotted line (plot (c) only) is for the case where the ions have been charged balanced by adding H^+ . Inset: the three contributions to the total calculated *GF*. Dots – measured values for the SX extracts; dotted line – the SX extracts plus the calculated contribution of the individual polar organic compounds); solid line – the previous two contributions plus that calculated for the inorganic ions (in this case for mean of the two cases shown in the main plot).

ACCEPTED MANUSCRIPT

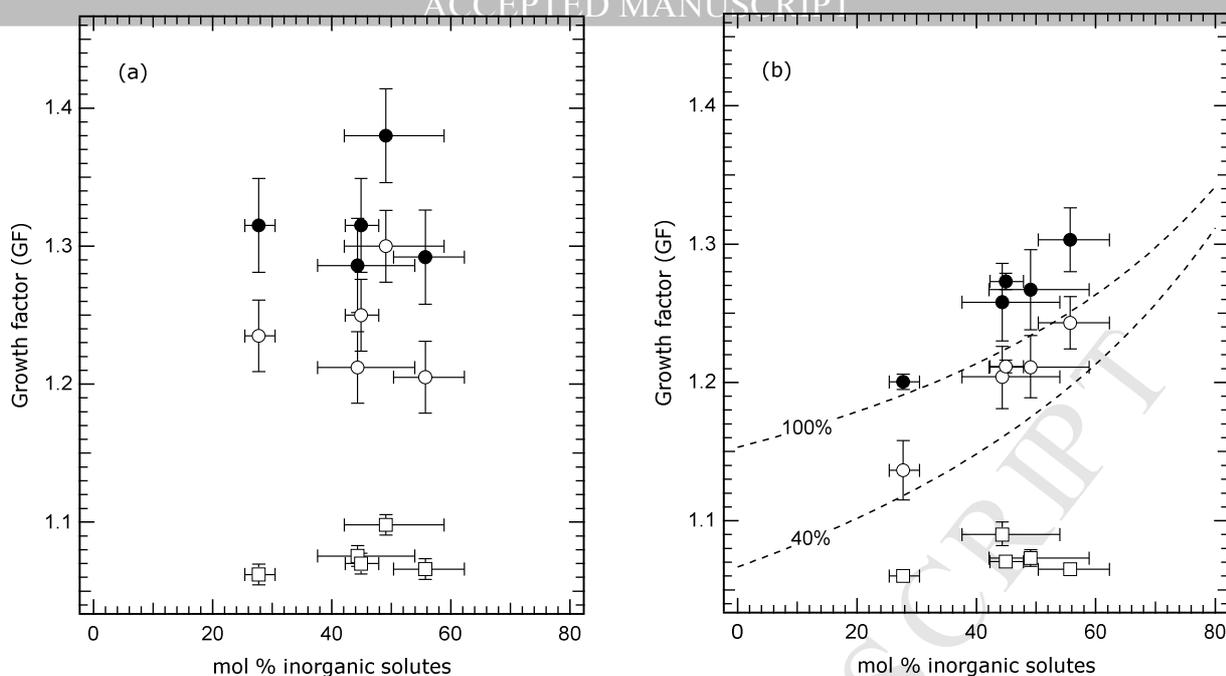


Figure 9. Growth factors (GF) for all samples except S5, at three relative humidities, plotted against the mol % of inorganic solutes present in the samples. (a) Measured values; (b) calculated (the same as in Figure 8). Symbols: open squares – 60% RH , open circles – 80% RH ; dots – 85% RH . Lines: calculated values, for 80% RH , for a mixture of $(NH_4)_2SO_4$ and a "Raoult's law" organic compound having the same average molar volume as the WSOC material in samples SX1-SX6, and the same assumed solid density (1.3 g cm^{-3}). Calculated GF for the upper line assume that 100% of the organic is dissolved in the aqueous phase, and 40% dissolved for the lower line. The bars associated with each value represent the following: horizontal – the range from the minimum mol % of inorganic solutes (in which ion amounts are adjusted to achieve charge balance to the lower of the cation and anion charge amounts), to the maximum (where the adjustment is to the higher of the two total charge amounts); vertical – the change in measured growth factor corresponding to an uncertainty of $\pm 2\%$ in RH .

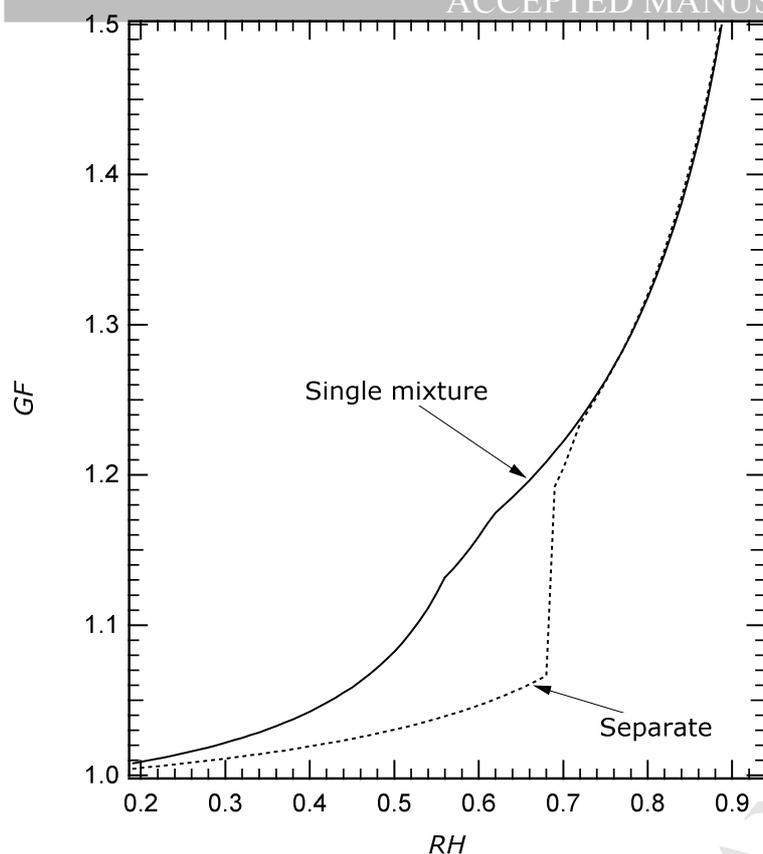


Figure 10. Calculated growth factors (GF) as a function of relative humidity (RH) for an aerosol consisting of 1 mole of a soluble organic compound (with water uptake conforming to Raoult's law), and 1 mole of $(\text{NH}_4)_{1.5}\text{H}_{0.5}\text{SO}_4$ (letovicite). Lines: solid – the mixture is modelled as a single solution, containing both solutes; dotted – the mixture is modelled according to eq (5), as the sum of the volumes of one solution containing only the organic compound and water, and a second solution containing the letovicite and water (including solid salts that form at low RH).

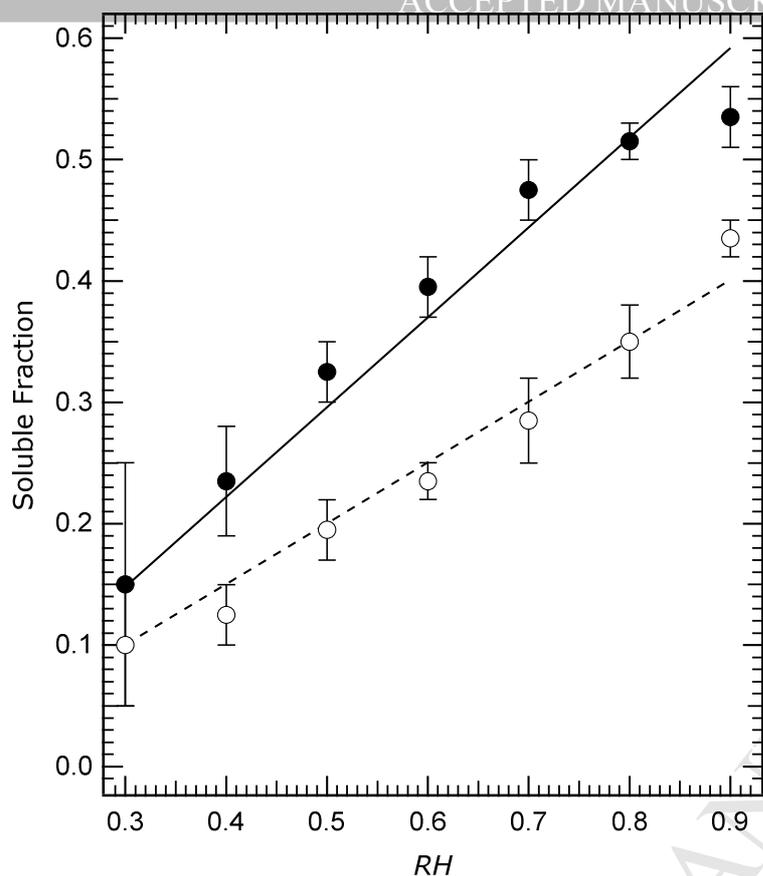


Figure 11. Soluble fractions of the WSOC material for which calculated growth factors agree with the measured values, plotted against relative humidity (RH). Symbols: dots – values for the assumption of Raoult's law mixing in the aqueous phase (Figure 7d); open circles – values from the UNIFAC calculation of equilibrium RH in which the functional group assignments gave high weight to alkane, $-OH$, and $-COOH$ functional groups (Figure 7c). Lines: solid – fitted line for the Raoult's law case, given by $0.74(RH - 0.1)$; dashed – fitted line for the UNIFAC case, given by $0.50(RH - 0.1)$. Note: soluble fractions in the figure are mole-based, and can be converted to mass by multiplying by 0.907. The bars associated with each point represent the soluble fractions associated with the upper and lower measured growth factors, and do not represent experimental error.

- A hygroscopicity index is developed for secondary organic aerosol material
- The estimated molecular group compositions confirm large numbers of –OH and –COOH
- Dissolution of water-extractable organic material varies linearly with RH
- Results suggest simple methods of modelling water uptake of atmospheric organics
- Measured and predicted growth factors of the total soluble aerosol agree well

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Completed by Simon Clegg on behalf of all authors.