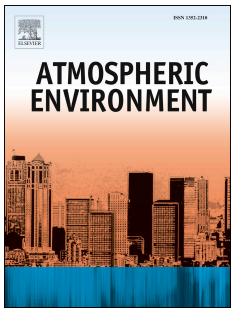
# Accepted Manuscript




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.

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

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

[a] *School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *Air Quality Research Center, University of California, Davis, California, USA*
[c] *Atmospheric Science Program, Michigan Technological University, Houghton, Michigan, US*
[d] *Department of Chemistry, Michigan Technological University, Houghton, Michigan, USA*
[e] *Division of Atmospheric Sciences, Desert Research Institute, Reno, Nevada, USA*
[f] *Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA*
[g] *Department of Chemical and Environmental Engineering, University of California, Riverside, California, USA*
[h] *Department of Atmospheric Science, University of Utah, Salt Lake City, Utah, USA*

## ABSTRACT

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

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

## 1. Introduction

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

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

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

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

## 2. Data

The sampling of atmospheric aerosols was performed from 24 June to 28 July 2010 at the Storm Peak Laboratory (3210 m above sea level, ASL), which is a remote continental site near Steamboat Springs (Colorado, USA). These aerosols are likely to be typical of many remote locations dominated by biogenic aerosol formation and the results of this study are likely to apply to similar aerosols elsewhere. The sampling protocol, treatment of the aerosol samples, and the chemical analyses are described by Hallar et al. (2013). Briefly, the samples were collected on two types of filters: (i) Teflon impregnated glass fiber filters (TIGF, filter size 8" × 10", Fibrefilm T60A20, PALL, Port Washington, NY) for sampling of $PM_{2.5}$ at a flow rate of ~ 1 $m^3$ $min^{-1}$, and (ii) pre-fired quartz-fiber filters (filter size: 47 mm; 2500 Pallflex QAT-UP, PALL, Port Washington, NY) for sampling of aerosols at a flow rate of ~0.11 $m^3$ $min^{-1}$. Daily filter samples were combined into six composites (the S1 to S6 that are the subject of this study) based on meteorological conditions and backward trajectories (Hallar et al., 2013). Material from the TIGF filters was used for the analysis of inorganic ions, individual water-soluble organic compounds (Samburova et al., 2013), molecular formula characterization of the higher molecular weight water-soluble organic fraction (Mazzoleni et al., 2012), and hygroscopicity measurements of water extracts with a TDMA (Taylor et al., 2017). These hygroscopicities include those of both the total water-soluble aerosol material (containing the inorganic ions), and measurements for the high molecular weight water soluble organic matter only. Quartz fiber filters were used for analysis of bulk elemental carbon (EC), organic carbon (OC), and water-soluble organic carbon (WSOC) (Hallar et al., 2013; Samburova et al., 2013).

All samples were analysed for the inorganic ions $Na^+$, $K^+$, $Mg^{2+}$, $Ca^{2+}$, $SO_4^{2-}$, $NO_3^-$ and $Cl^-$ by ion chromatography and automated colourimetry ($NH_4^+$ only) (Samburova et al., 2013); and for individual polar organic species (acids, sugars, sugar alcohols, sugar anhydrates, and lignin derivatives) by a combination of IC and GC-MS (Samburova et al., 2013). The extraction of the other water-soluble organic matter (WSOC), using XAD-8 and XAD-4 resins, is also described by Samburova et al. (2013). The WSOC compounds were characterised by ultrahigh resolution Fourier transform-ion cyclotron resonance MS (FT-ICR MS) (see Mazzoleni et al., 2012). The composite samples are referred to in this work as S1-S6 (total water-soluble aerosol material), and SX1-SX6 (water-soluble high molecular weight organic matter only). A schematic diagram of the chemical analyses carried out on the samples is shown in Figure 1 of Hallar et al. (2013).

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

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

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

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

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

*2.1. Inorganic ions*

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

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

*2.2 Polar organic compounds*

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

*2.3 WSOC compounds*

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

In addition to the full scan analysis, FT-ICR MS/MS fragmentation analysis was used to investigate the functional groups present in the identified molecular formulae in sample extract SX4. Due to the extreme isobaric complexity of water-soluble organic aerosol, individual mass spectral peaks could not be isolated for fragmentation. Instead small mass range windows (6 or 10 u) were selected for fragmentation, consistent with LeClair et al. (2012). Each mass window was defined by a central mass selected at intervals of 5 u (m/z 180, 185, etc.) over the range of m/z 160 - 365 and then every 10 u over the range of m/z 365-485. Ultrahigh resolution analysis using FT-ICR MS was done on both unfragmented ions (representing precursor ions) and the fragmented ions after collision induced dissociation (representing product ions). Molecular formulae were then assigned to the collected ultrahigh resolution mass spectra using Composer software as described in Mazzoleni et al. (2012). The resulting precursor and fragment formulas were paired based on the exact mass differences associated with expected common neutral losses (e.g., $CO_2$, $H_2O$, etc.). A total of 1471 precursor formulas were assigned to the studied mass ranges and 100% of them were also found in the full scan analysis of this same sample reported in Mazzoleni et al. (2012).

Quantitative information regarding the molecules in this component of the aerosol is limited to the amounts, numbers of C, H, O, N, and S atoms in each molecule, and the numbers of double bond equivalents (DBE). 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 three elements in the molecule. Note that one ring counts as 1 DBE, a triple bond counts as 2 DBE and an aromatic ring is 4 DBE (one for the ring plus one for each C=C). It is polar groups such as –OH and –COOH that particularly influence solubility in water and the relationship between water activity (equilibrium *RH*) and concentration or hygroscopicity. The results of the FT-ICR MS analysis give some insight into the abundances of the different functional groups in the WSOC material for sample SX-4, in the following way. A total of 21 different neutral losses were observed for the studied precursor and fragment ion molecular formulas based on exact mass difference pairing. For example, a neutral loss of $H_2O$ is indicative of a hydroxyl functional group (–OH) and a neutral loss of $CO_2$ is indicative of a carboxyl functional group (–COOH). Likewise, a neutral loss of $CH_2O_3$ is indicative of two functional groups (carboxyl (–COOH) and hydroxyl). Combinations of neutral losses are expected for multifunctional

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

The neutral losses were grouped into five major categories: $CO_2$ losses, $H_2O$ losses, methoxy losses, aldehyde losses, and nitrogen and/or sulfur losses. Some neutral losses can fit into two categories, for example the $CH_2O_3$ loss mentioned previously is classified as both a $CO_2$ loss and an $H_2O$ loss because both functional groups are contained within that neutral loss. This means that some losses will be counted twice, once in two different categories. The two most abundant loss categories were $CO_2$ and $H_2O$, which were observed for 1279 (86.9%) and 1339 (91.0%) of the precursor formulas overall. The two next most abundant were aldehyde (1148, 78.0%) and methoxy (978, 66.5%) group neutral losses. The complete breakdown of this is shown in Table 5, and the results are discussed further in the Appendix.

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

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

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

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

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

### *3.1 Hygroscopicity index*

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

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

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

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

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

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

*3.2 Categorising the compounds*

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

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

*3.3 Varying the assignment of functional groups*

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

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

## 4. Methods

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

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

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

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

$$\Sigma_i (m_i / m_i^o) = 1 \quad (3)$$

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

$$W_T = \Sigma_i \, w_i^o \quad (4)$$

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

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

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

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

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

where $V_T$ is the total volume of the aqueous mixture at water activity $a_w$, and $V^o$ 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):

$$1/\rho = \Sigma_i x_i^* / \rho_i^o \quad (7)$$

where $\rho$ is the density of the mixture, and $\rho_i^o$ 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:

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

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 $V^o$ in equation 6) is related to its density by $V^o = M_T / \rho$ 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 $H_{0.5}SO_4^{1.5-}$ 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 ($\Sigma_c n_c z_c$ and $\Sigma_a n_a|z_a|$) 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, $CaSO_4.2H_2O$) is very small (about 0.015 mol $kg^{-1}$ at 25 $^oC$) 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

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

*4.2 Polar organic compounds*

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

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

*4.3 WSOC fraction*

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

## 5. Modelling Hygroscopic Growth Factors of the Organic Material

It is well understood that electrolyte solutes are more hygroscopic, and have higher growth factors, than most soluble organic compounds. This is illustrated in Figure 5, which compares measured growth factors of the SX1 WSOC material, calculated growth factors of the S1 polar organics, and calculated growth factors of $(NH_4)_2SO_4$, $(NH_4)_3H(SO_4)_2$, and $NH_4HSO_4$. The polar organic fraction was assumed to be fully liquid at all *RH* (hence the continuous increase in GF with *RH*, in contrast to the deliquescence transitions shown for the salts). The growth factor would be increased by a factor of about 1.057 if the polar organic fraction were assumed to be solid at the reference *RH* (this is based upon the difference between the solid and liquid molar volumes of glucose, which is a significant component of the polar organic fraction). The measured growth factors of the WSOC material in Figure 5 are very low, which is also true of the other composite samples (see Figure 5 of Hallar et al., 2013). A simple calculation suggests that, at 80% *RH*, about 840 g of WSOC material is required to take up the same amount of water as 1 mole (132 g) of ammonium sulphate. Figure 1a shows that a comparable mass ratio of WSOC material to inorganic electrolytes (840 : 132, or 6.3 : 1) is approached only in sample S2. In the other samples, the ratio varies from about 1:1 (S1) to about 2.5:1 (S5).

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

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

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

*5.1 Polar organic compounds*

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

*5.2 The WSOC fraction*

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

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

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

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

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

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

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

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

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

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

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

First, alculated growth factors are generally lower than measured values. This is especially the case for sample S2, which has the lowest inorganic fraction of all the samples (see Figure 1a,c) and consequently the lowest predicted growth factor. In Figure 9 we show the measured growth factors for all samples at three selected *RH*, plotted against the mol % of inorganic solutes. There is no apparent relationship between the two, which is not what would be expected. We investigated this behaviour further by plotting the same quantities, but using the calculated growth factors, in Figure 9b. To this we added the calculated growth factors, for 80% *RH*, of a mixture of 1 mole of $(NH_4)_2SO_4$ and 1 mole of organic material, which is assumed to take up water according to Raoult's law and has a dry density and molar volume (when dissolved) the same as the WSOC material. The two dashed lines on the plot correspond to 40% dissolved organic material, and 100%. The results in Figure 9b show, first, that there is essentially no relationship expected between mole % of inorganic solutes and growth factor at 60% *RH*, because a significant fraction is calculated to be solid at this point. This is in agreement with the measured values for the same *RH*, shown in Figure 9a. At 80% *RH* the calculated *GF* for the five sample compositions show a relationship with inorganic content with a slope that corresponds quite closely to the two simplified cases (the dashed lines in the Figure 9b). However, this is not the case for the measured growth factors. The main factors that affect these comparisons are: (i) uncertainties in the TDMA measurements, (ii) uncertainties in the composition measurements and the

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

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

Second, there are some differences between the growth factors measured in the two scans for each sample (particularly S4), but in general they are slight. The calculated growth factors also do not show the deliquescence "steps" that are typical of simple purely inorganic systems, except for sample S1. For this sample the increased growth factor at about 72% *RH* is due mostly to the predicted dissolution of $(NH_4)_2SO_4$ and $MgSO_4.6H_2O$. In all samples numerous salts are predicted to be formed, and the smoothness of most of the growth factor curves in Figure 8 can be attributed to the formation and dissolution of large numbers of salts as *RH* changes. For example, for sample S3, over the *RH* range 80% to 20%, the following solid salts are present: $K_2SO_4$, $MgSO_4.6H_2O$, $Na_2SO_4.(NH_4)_2SO_4.4H_2O$, $(NH_4)_2SO_4$, $\{2,3\}NH_4NO_3.(NH_4)_2SO_4$, $NH_4Cl$, and $MgSO_4.H_2O$.

Third, it is important to remember that the calculation of the water content of the samples as the sum of the three different components (equation 5) introduces an artifact with regard to the formation of the solid salts. This is illustrated in Figure 10 for the case of a hypothetical aerosol containing 1 mole of a soluble "Raoult's law" organic compound and one mole of $(NH_4)_{1.5}H_{0.5}SO_4$ (letovicite). Where the water content associated with the organic compound and salt are calculated separately, the electrolyte fraction contributes nothing to the total water content below the deliquescence point (68% *RH*), whereas in reality the soluble organic fraction of the aerosol provides some water at all *RH* for the salt to dissolve into. This leads to much higher growth factors at moderate to low *RH*, a smoother growth factor curve, and a reduction (i.e., lower *RH*) of the deliquescence transitions with respect to $(NH_4)_2SO_4$ and $(NH_4)_{1.5}H_{0.5}SO_4$. We would not expect such a large effect for the measured samples, because of the apparently small amounts of water associated with the organic fraction, and it would not explain the difference between measured and modelled growth factors at high *RH* for which there are no solid salts.

Perhaps more relevant to the present study is the fact that, at high *RH*, equation (5) does not take into account the influence of the additional amounts of WSOC organic material that can be expected to dissolve into the relatively large amount of water associated with the inorganic fraction of the total aerosol. There are two elements to consider: first, the additional volume of water for the organic to dissolve in; second, the change in the activity coefficient ($\gamma$) of the organic going from the solution of WSOC material + water (as measured for the SX series of extracts) to a mixture that also contains the inorganic solutes. The latter is probably of lesser importance: a calculation using the Zdanovskii-Stokes-Robinson expression for solute activity coefficients (equation 9 of Clegg et al., 2003) suggests a change from $\gamma$(WSOC) equal to 0.8 (on a molality basis) in a water-WSOC solution at 80% RH, to 0.66 in a mixture containing 1 mole of dissolved $(NH_4)_2SO_4$ and 0.5 moles of dissolved WSOC material. In this calculation the ammonium sulphate is used as a surrogate of the more complex inorganic mixtures occurring in the samples, and the 0.5 moles of WSOC material corresponds to a 1:1 mixture (in terms of moles) in which 50% dissolves in water. (The mole % compositions of the samples can be seen in Figure 1c.)

The *E-AIM* model was used to investigate the effect of additional dissolution of WSOC organic material in the total aerosol samples (S1-S6), relative to that estimated to occur for SX1 in Figure 7 under various assumptions, as follows. The amounts of water associated with 1 mole of $(NH_4)_2SO_4$ and 1 mole of WSOC material (50% dissolved) at 80% *RH* are 9.56 and 2.0 moles, respectively. If we assume a WSOC "dry" density of 1.3 g $cm^{-3}$, a molar volume of 386 $cm^3$ $mol^{-1}$ when dissolved (for an average molar mass of 386 g $mol^{-1}$) then the growth factor of the mixture at 80% is 1.19. Also needed for this calculation is the density of dry $(NH_4)_2SO_4$, which is 1.77 g $cm^{-3}$, and that of its aqueous solution in equilibrium with 80% *RH* which is 1.245 g $cm^{-3}$. This *GF* corresponds quite closely with those predicted for samples S3-S6, which are the ones that are nearest to a 1:1 mixture of inorganic solutes and WSOC material. If a further 25% of the WSOC material dissolves into the water associated with the salt, we calculate in increase in the growth factor of 0.03 (to 1.22). This increase, which would bring the predicted *GF* closer to the measured ones for these total samples is significant and is of about the same magnitude as the uncertainty associated with the charge imbalances of the inorganic ion amounts for samples S3 and S4.

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

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

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

## 7. Discussion and Summary

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Acknowledgements**

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

# APPENDIX

## Estimation of the Functional Group Compositions

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

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

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

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

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

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

## REFERENCES


K. Balslev and J. Abildskov (2002) UNIFAC parameters for four new groups. *Ind. Eng. Chem. Res.* **41**, 2047-205.

O. Boucher, D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang (2013) Clouds and Aerosols. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, eds.T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge University Press.

C. M. Carrico, M. D. Petters, S. M. Kreidenweis, J. L. Collett Jr., G. Engling, and W. C. Malm (2008) Aerosol hygroscopicity and cloud droplet activation of extracts of filters from biomass burning experiments. *J. Geophys. Res.* **113**, D08206, doi:10.1029/2007JD009274.

S. L. Clegg, P. Brimblecombe, and A. S. Wexler (1998) A thermodynamic model of the system $H^+$ - $NH_4^+$ - $SO_4^{2-}$- $NO_3^-$ - $H_2O$ at tropospheric temperatures. *J. Phys. Chem. A* **102**, 2137-2154.

S. L. Clegg and J. H. Seinfeld (2006a) Thermodynamic models of aqueous solutions containing inorganic electrolytes and dicarboxylic acids at 298.15 K. II. Systems including dissociation equilibria. *J. Phys. Chem. A* **110**, 5718-5734.

S. L. Clegg and J. H. Seinfeld (2006b) Thermodynamic models of aqueous solutions containing inorganic electrolytes and dicarboxylic acids at 298.15 K. I. The acids as non-dissociating components. *J. Phys. Chem. A* **110**, 5692-5717.

S. L. Clegg, J. H. Seinfeld and P. Brimblecombe (2001) Thermodynamic modelling of aqueous aerosols containing electrolytes and dissolved organic compounds. *J. Aerosol. Sci.* **32**, 713-738.

S. L. Clegg, J. H. Seinfeld, and E. O. Edney (2003) Thermodynamic modelling of aqueous aerosols containing electrolytes and dissolved organic compounds. II. An extended Zdanovskii-Stokes-Robinson approach. *J. Aerosol. Sci.* **34**, 667-690.

S. L. Clegg and J. M. Simonson (2001) A BET model of the thermodynamics of aqueous multicomponent solutions at extreme concentration. *J. Chem. Thermodyn.* **33**, 1457-1472.

S. L. Clegg and A. S. Wexler (2011) Densities and apparent molar volumes of atmospherically important 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 $NH_4Cl$ from 0 to 50 $^oC$, including extrapolations to very low temperature and to the pure liquid state, and $NaHSO_4$, NaOH, and $NH_3$ at 25 $^oC$. *J. Phys. Chem. A* **115**, 3393-3460.

A. Fredenslund, R. L. Jones, J. M. Prausnitz, (1975) Group-contribution estimation of activity coefficients in non-ideal liquid mixtures AIChEJ., **21,** 1086-1099.

Freedman, M. A. (2017) Phase separation in organic aerosol. *Chemical Society Reviews*, **46**(24), 7694-7705.

G. S. Girolami (1994) A Simple "Back of the Envelope" Method for Estimating the Densities and Molecular Volumes of Liquids and Solids, *J. Chem. Educ.* **71**, 962-964.

A. G. Hallar, D. H. Lowenthal, S. L. Clegg, V. Samburova, N. Taylor, L. R. Mazzoleni, B. K. Zielinska, T. B. Kristensen, G. Chirokova, I. B. McCubbin, C. Dodson, D. Collins (2013) Chemical and hygroscopic properties of aerosol organics at Storm Peak Laboratory. *J. Geophys. Res.(Atmospheres)* **118**, 1-13, doi:10.1002/jgrd.50373.

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.

E. Jenkin, J. L. Jimenez, A. Kiendler-Scharr, W. Maenhaut, G. McFiggans, Th. F. Mentel, A. Monod, A. S. H. Prevot, J. H. Seinfeld, J. D. Surratt, R. Szmigielski, and J. Wildt (2009) The formation, properties and impact of secondary organic aerosol: current and emerging issues. *Atmos. Chem. Phys.* **9**, 5155-5236.

H. K. Hansen, P. Rasmussen, A. Fredenslund, M. Schiller, and J Gmehling (1991) Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **30**, 2352-2355.

Y.-F. Hu (2000) An empirical approach for estimating the density of multicomponent aqueous solutions obeying the linear isopiestic relation, *J. Solut. Chem.* **29**, 1229-1236.

Mark Z. Jacobson (1999) *Fundamentals of Atmospheric Modelling*, Cambridge University Press, 656 pp.

M. Kanakidou, J. H. Seinfeld, S. N. Pandis, I. Barnes, F. J. Dentener, M. C. Facchini, R. Van Dingenen, B. Ervens, A. Nenes, C. J. Nielsen, E. Swietlicki, J. P. Putaud, Y. Balkanski, S. Fuzzi, J. Horth, G. K. Moortgat, R. Winterhalter, C. E. L. Myhre, K. Tsigaridis, E. Vignati, E. G. Stephanou, and J.Wilson (2005) Organic aerosol and global climate modelling: a review, *Atmos. Phys. Chem.* **5**, 1053-1123.

B. P. Koch and T. Dittmar (2006) From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter, *Rapid Communications in Mass Spectrometry* **20**, 926-932. See also the erratum published in 2016 in volume 30, page 250, of the same journal.

T. B. Kristensen, H. Wex, B. Nekat, J. K. Nøjgaard, D. van Pinxteren, D. H. Lowenthal, L. R. Mazzoleni, K. Dieckmann, C. B. Koch, T. F. Mentel, H. Herrmann, A. G. Hallar, F. Stratmann, and M. Bilde (2012) Hygroscopic growth and CCN activity of HULIS from different environments, *J. Geophys. Res.* **117**, D22203, doi:10.1029/2012JD018249.

J. P. LeClair, J. L. Collett, and L. R. Mazzoleni (2012) Fragmentation analysis of water-soluble atmospheric organic matter using ultrahigh-resolution FT-ICR mass spectrometry. *ES&T* **46**, 4312-4322.

Numerical Algorithms Group (2013) *The NAG Fortran Library, Mk. 24*, Oxford, (http://www.nag.co.uk).

S. T. Martin (2000) Phase transitions of aqueous atmospheric particles. *Chem. Rev.* **100**, 3403-3453.

L. R. Mazzoleni, P. Saranjampour, M. M. Dalbec, V. Samburova, A. G. Hallar, B. Zielinska, D. H. Lowenthal, and S. D. Kohl (2012) Identification of water-soluble organic carbon in non-urban aerosols using ultrahigh-resolution FT-ICR mass spectrometry: organic anions, *Environ. Chem.* **9**, 285-297.

M. D. Petters and S. M. Kreidenweis (2007) A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, **7**, 1961–1971.

M. D. Petters and S. M. Kreidenweis (2008) A single parameter representation of hygroscopic growth and cloud condensation nucleus activity – Part 2: Including solubility, *Atmos. Chem. Phys.*, **8**, 6273–6279.

C. A. Pope III, and D. W. Dockery (2006) Health effects of fine particulate air pollution: lines that connect, *J. Air Waste Manage.*, **56**, 709–742.

I. Riipinen, N. Rastak, and S. N. Pandis (2015) Connecting the solubility and CCN activation of complex organic aerosols: a theoretical study using solubility distributions, *Atmos. Chem. Phys.* **15**, 6305-6322.

V. Samburova, A. G. Hallar, L. R. Mazzoleni, P. Saranjampour, D. H. Lowenthal, S. D. Kohl, and B. Zielinska (2013) Composition of water-soluble organic carbon in non-urban atmospheric aerosol collected at the Storm Peak Laboratory, *Environ. Chem.* **10**, 370-380.

M. Semmler, B. P. Luo, T. Koop (2006) Densities of liquid $H^+$ / $NH_4^+$ / $SO_4^{2-}$ / $NO_3^-$ / $H_2O$ solutions at tropospheric temperatures. *Atmos. Environ.*, **40**, 467–483.

J. H. Seinfeld and S. N. Pandis (2006) *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd Edn. (revised), Wiley-Blackwell, 1232 pp.

M. Shiraiwa, Y. Li, A. P. Tsimpidi, V. A. Karydis, T. Berkemeier, S. N. Pandis, J. Lelieveld, T. Koop, and U. Poschl (2017) Global distribution of particle phase state in atmospheric secondary organic aerosols. *Nature Communications*, **8**, 15002 (doi: 10.1038/ncomms15002).

J. D. Smith, V. Sio, L. Yu, Q. Zhang, and C. Anastasio (2014) Secondary organic aerosol production from aqueous reactions of atmospheric phenols with an organic triplet excited state. *ES&T* **48**, 1049-1057.

R. H. Stokes, and R. A. Robinson (1966) Interactions in aqueous nonelectrolyte solutions: I. Solute-solvent equilibria. *J. Phys. Chem.* **70**, 2126-2130.

S. Suda Petters, D. Pagonis, M. S. Claflin, E. J. T. Levin, M. D. Petters, P. J. Ziemann, and S. M. Kreidenweis (2017) Hygroscopicity of organic compounds as a function of carbon chain length and carboxyl, hydroperoxy, and carbonyl functional groups. *J. Phys. Chem. A* **121**, 5164-5174.

S. Suda Petters, S. M. Kreidenweis, A. P. Grieshop, P. J. Ziemann, and M. D. Petters (2019) Temperature- and humidity-dependent phase states of secondary organic aerosols. *Geophys. Res. Lett.*, doi:10.1029/2018GL080563

N. F. Taylor, D. R. Collins, D. H. Lowenthal, I. B. McCubbin, A. G. Hallar, V. Samburova, B. Zielinska, N. Kumar, and L. R. Mazzoleni (2017) Hygroscopic growth of water soluble organic carbon isolated from atmospheric aerosol collected at US national parks and Storm Peak Laboratory. *Atmos. Chem. Phys.* **17**, 2555-2571.

C. H. Tong, S. L. Clegg and J. H. Seinfeld (2008) Comparison of activity coefficient models for atmospheric aerosols containing mixtures of electrolytes, organics, and water. *Atmospheric Environment* **42**, 5459-5482.

T. V. Vu, J. M. Delgado-Saborit, and R. M. Harrison (2015) A review of hygroscopic growth factors of submicron aerosols from different sources and its implication for calculation of lung deposition efficiency of ambient aerosols. *Air Qual. Atmos. Health* **8**, 429-440.

J. Wang, M. J. Cubison, A. C. Aiken, J. L. Jimenez, and D. R. Collins (2010) The importance of aerosol mixing state and size-resolved composition on CCN concentration and the variation of the importance with atmospheric aging of aerosols. *Atmos. Chem. Phys.* **10**, 7267-7283.

A. S. Wexler and S. L. Clegg (2002) Atmospheric aerosol models for systems including the ions $H^+$, $NH_4^+$, $Na^+$, $SO_4^{2-}$, $NO_3^-$, $Cl^-$, $Br^-$, and $H_2O$. *J. Geophys. Res.-Atmos.* **107**, D14, Art. No. 4207, (see http://www.aim.env.uea.ac.uk/aim/aim.php).

R. Wittig, J. Lohmann, and J. Gmehling (2003) Vapor-Liquid Equilibria by UNIFAC Group Contribution. 6. Revision and Extension. *Ind. Eng. Chem. Res.* **42**, 183-188.

M. Zark, J. Christophers, and T. Dittmar (2017) Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar. Chem.* **191**, 9-15.

R. A. Zaveri, R. C. Easter, J. D. Fast, and L. K. Peters (2008) Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), *J. Geophys. Res.* **113**, D13204, doi:10.1029/2007JD008782.

A. Zuend, C. Marcolli, B. P. Luo, and T. Peter (2008) A thermodynamic model of mixed organic-inorganic aerosols to predict activity coefficients. *Atmos. Chem. Phys.* **8**, 4559-4593.

# TABLES (main text)

**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 |
| **$NO_3^-$** | nmol $m^{-3}$ | 1.340 | 1.345 | 2.213 | 2.032 | 0.436 | 0.808 |
| **$SO_4^{2-}$** | nmol $m^{-3}$ | 5.662 | 3.435 | 4.709 | 0.805 | 1.008 | 2.518 |
| **$NH_4^+$** | 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 |
| **$Ca^{2+}$** | nmol $m^{-3}$ | 0.707 | 0.665 | 0.661 | 0.829 | 0.348 | 0.812 |
| **$Mg^{2+}$** | 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 $m^3$ of atmosphere sampled. The analytical uncertainties for inorganic ions were: 3-10 % ($Cl^-$), 1-23 % ($NO_3^-$), 2-16 % ($SO_4^{2-}$), 3-12 % ($NH_4^+$), 0.2-1 % ($Na^+$), 1-4 % ($K^+$), 1-5 % ($Ca^{2+}$), 4-14 % ($Mg^{2+}$).

[a] Charge balance is calculated as $(\Sigma_i\, nC_i\, zC_i - \Sigma_i\, nA_i\, |zA_i|)$ 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 $0.5(\Sigma_i\, nC_i\, zC_i + \Sigma_i\, nA_i\, |zA_i|)$.

[b] The charge balance is calculated on the assumption that $SO_4^{2-}$ is partially neutralised, and present in the aerosol as $HSO_4^-$.

[c] The charge balance is calculated on the assumption that $SO_4^{2-}$ is partially neutralised, and present in the aerosol as $H_{0.5}SO_4^{1.5-}$.

**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 |
| **$NO_3^-$** | nmol $m^{-3}$ | 0.0247 | 0.0432 | 0.0384 | 0.0376 | 0 | 0.0450 |
| **$SO_4^{2-}$** | nmol $m^{-3}$ | 0.0194 | 0.0462 | 0.0301 | 0.0187 | 0.0886 | 0.0251 |
| **$NH_4^+$** | 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 |
| **$Ca^{2+}$** | nmol $m^{-3}$ | 0.0609 | 0.0421 | 0.0671 | 0.0514 | 0 | 0.0582 |
| **$Mg^{2+}$** | 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 $m^3$ 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 $m^3$ 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 | $CO_2$ Loss (-COOH) | $H_2O$ 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.

# TABLES (appendix)

**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.

| | **Percentage Contributions of Each Type of Functional Group** | | |
|---|---|---|---|
| **Functional Group** [a] | **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 ($CH_2$); alkene - 2 (C=C); alcohol - 5 (OH); ketone - 9 ($CH_2CO$); aldehyde - 10 (CH); acid - 20 (COOH); oxide - 53 (see Balslev and Abildskov, 2002); nitro - 26 ($CNO_2$); amide - 46 (CON); thiol - 29 ($CH_3SH$); sulphide - 48 ($CH_2S$); sulphone - 55 (see Wittig et al., 2003); acetate - 11 (CCOO); ether - 13 ($CH_2O$). 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), $–CH_2–$, and $–CH_3$.

[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)

| | **Numbers of Each Type of Functional Group** | | |
|---|---|---|---|
| **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 | ΔS1 | ΔS2 | ΔS3 | ΔS4 | ΔS5 | ΔS6 |
|---|---|---|---|---|---|---|---|
| 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 ΔS1, ΔS2, etc. indicate the deviations of each of the six samples from that mean.

**FIGURES**

FIGURES

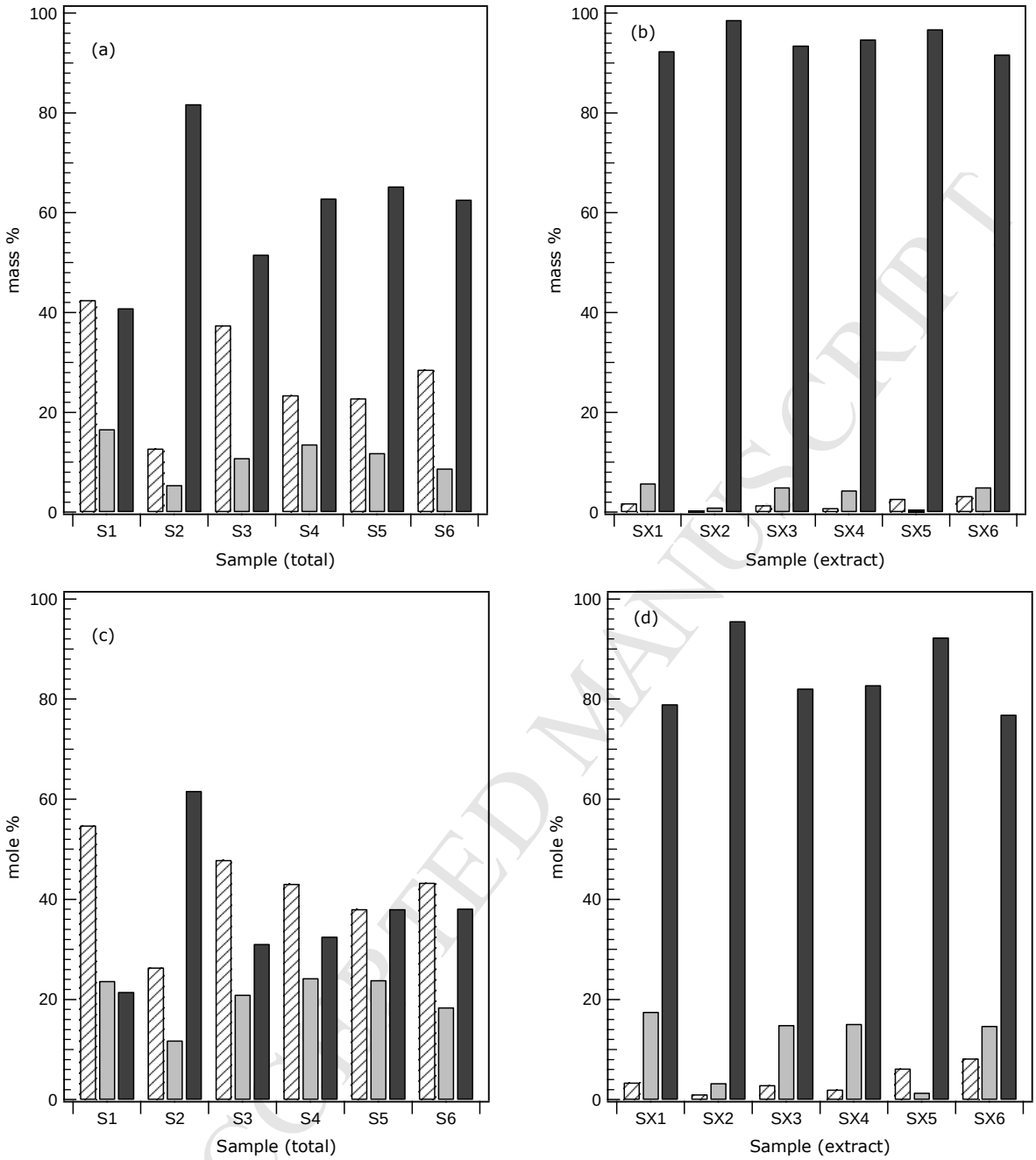



Figure 1. Relative compositions of the total water soluble aerosol material (samples S1-6), and water soluble organic matter extracts (SX1-6), in both mass % (a and b) and mole % (c and d). Key: diagonal lines – inorganic ions; solid grey – individual polar organic compounds; black – high molecular weight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

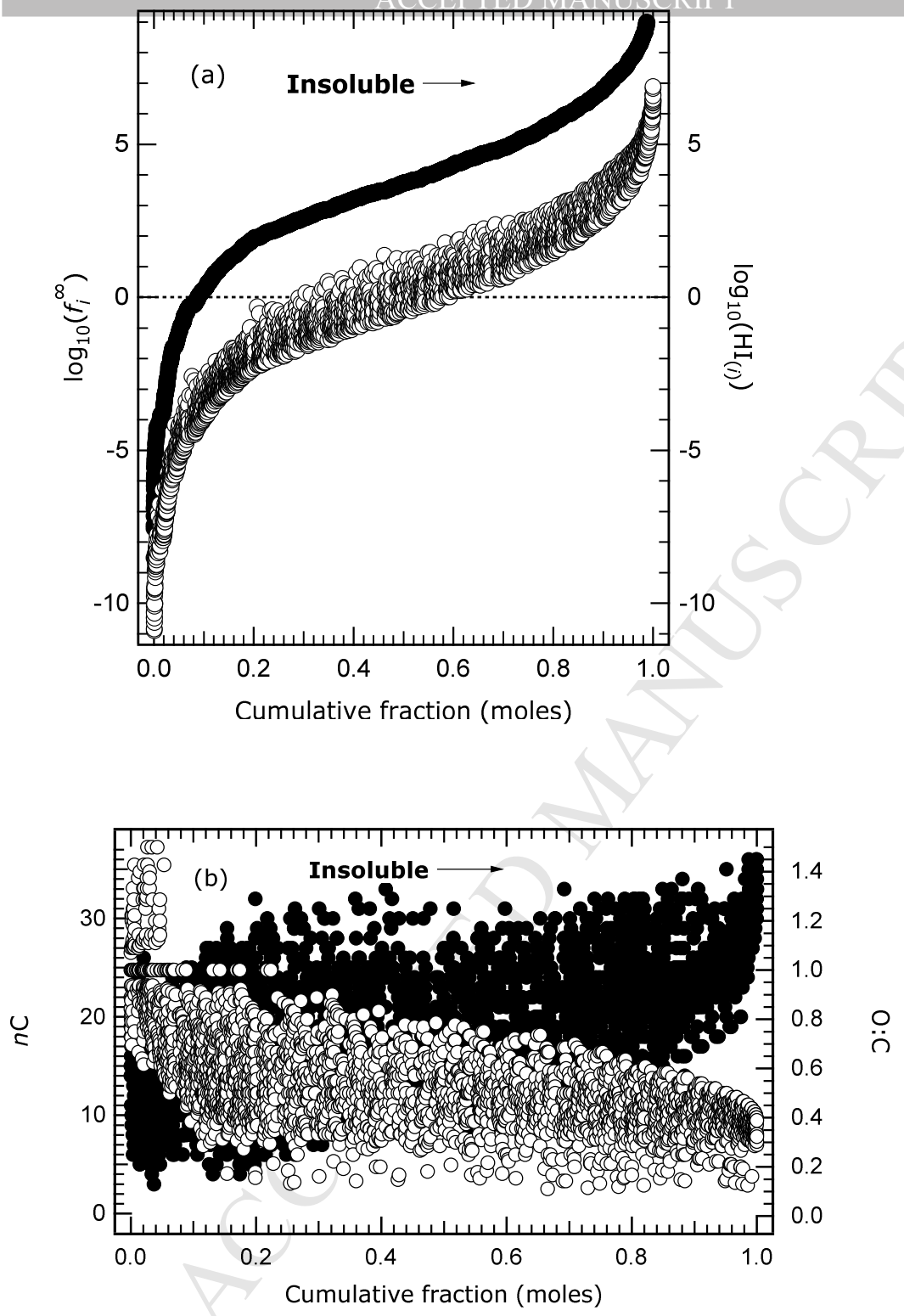



Figure 2. Calculated hygroscopicity index properties of WSOC sample SX1, plotted against the cumulative mole fraction of the sample, ranked in order of increasing values of the index *HI*. The most hygroscopic compounds are at the left of each plot, and the least hygroscopic compounds at the right. (a) Left axis and solid dots: the logarithm of the calculated activity coefficient of each compound at infinite dilution in water at 25 $^{o}$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.

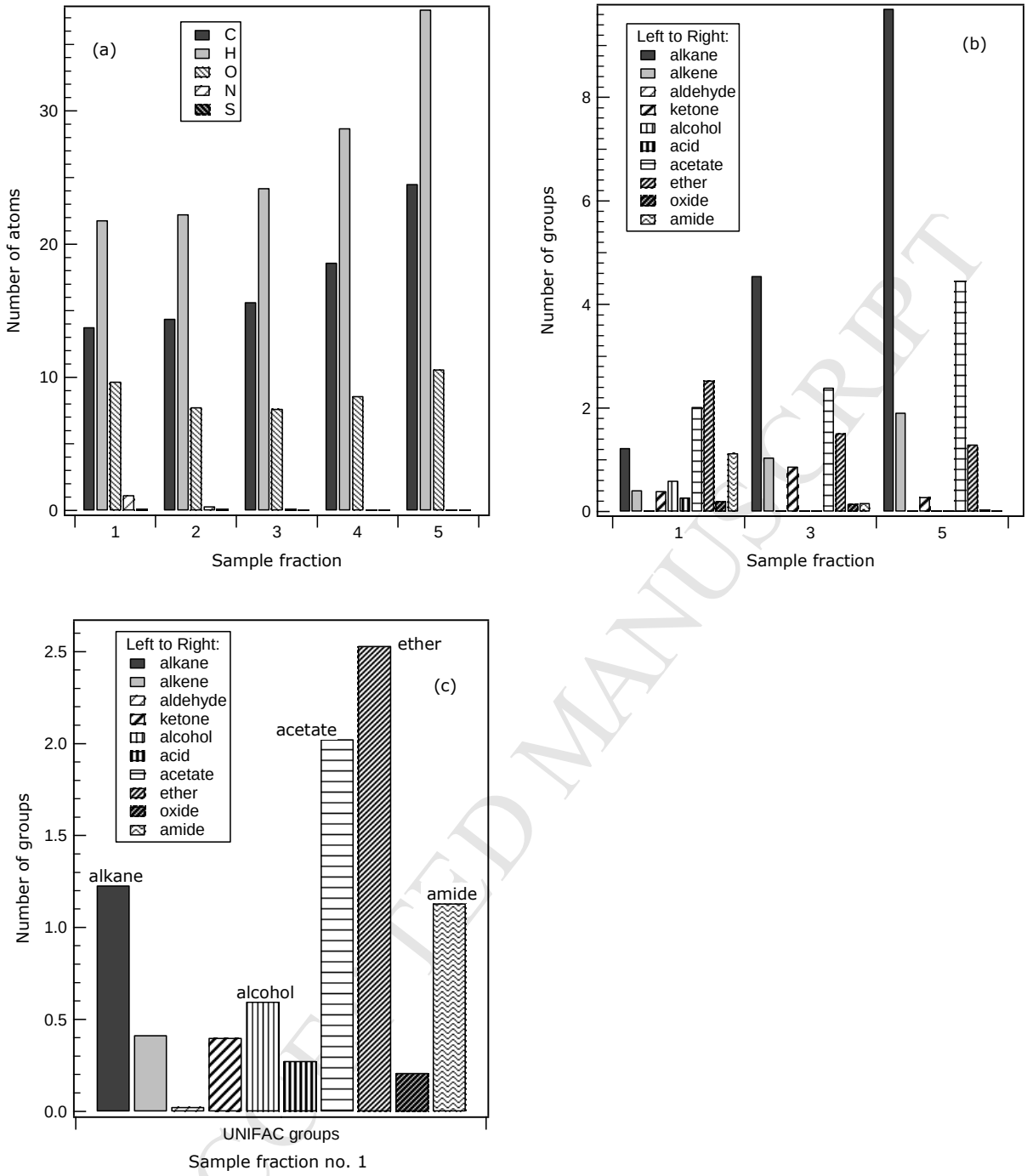



Figure 3. Characteristics of the WSOC fraction in sample SX1, with molecules ranked according to hygroscopicity index and then divided into five groups containing equal numbers of moles of each compound. This is the "base case" in which the functional group assignments are carried out so as to minimise 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, an 5. (c) Average UNIFAC group composition of fraction 1 (the same data as shown in b).

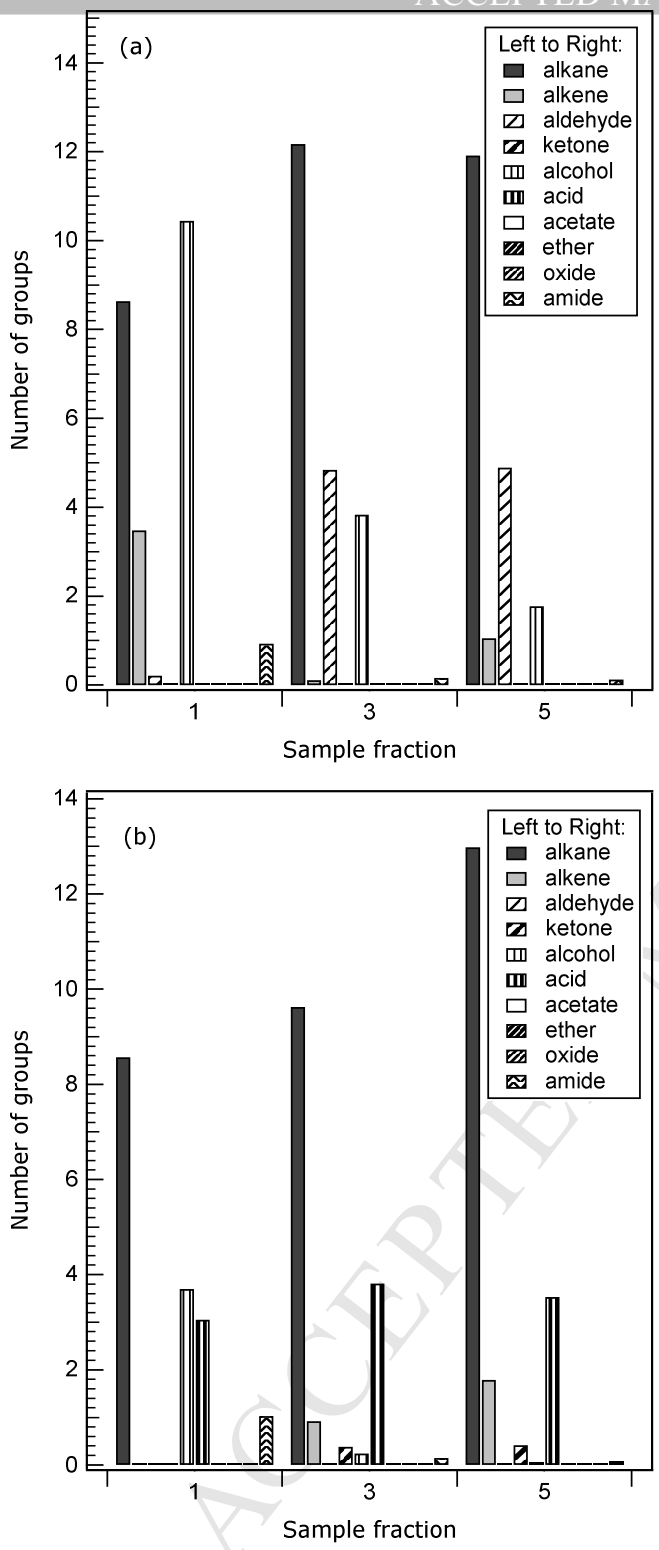



Figure 4. Predicted average UNIFAC group functional composition of the WSOC fraction in sample SX1, with molecules ranked according to hygroscopicity index *HI* and then divided into five groups containing equal numbers of moles of each compound. Fraction 1 is predicted to be the most soluble, and fraction 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.

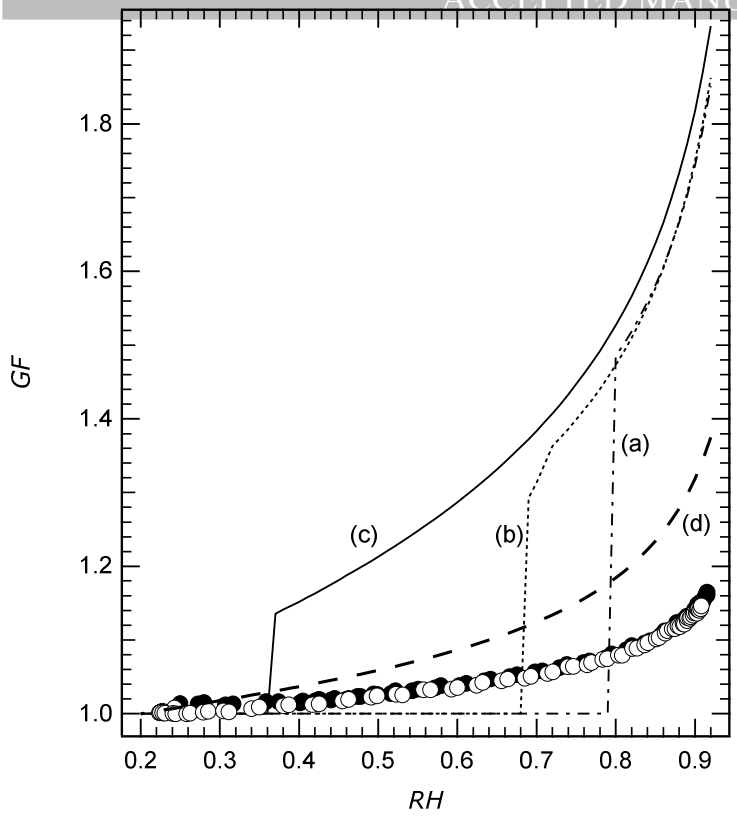



Figure 5. Calculated growth factors (*GF*) of three ammonium sulphate salts, and polar organic compounds, and measured growth factors of WSOC material SX1 plotted against relative humidity (*RH*). Symbols (WSOC material): dots – deliquescence (low *RH* to high *RH*); open circles – efflorescence (high *RH* to low *RH*). Lines: (a) $(NH_4)_2SO_4$; (b) $(NH_4)_3H(SO_4)_2$; (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-AIM* Model II of Clegg and Wexler (1998) for a temperature of 25 $^{o}$C, and for the polar organic compounds using UNIFAC, liquid molar volumes estimated using Girolami (1994), and equation (7) of this work.

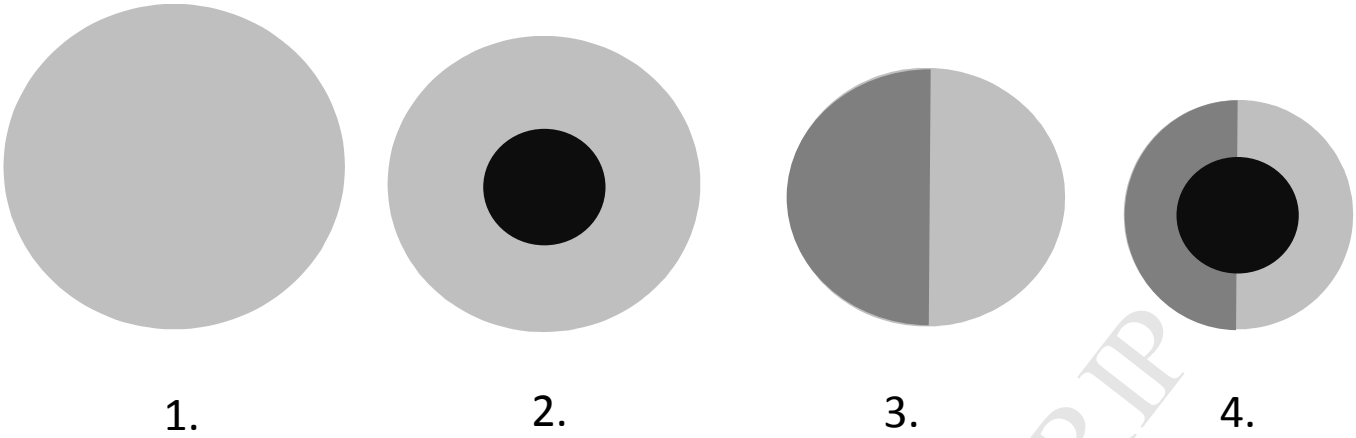



Figure 6. Schematic diagram of possible phases present in WSOC material that has taken up water. (1) A single liquid phase in which all WSOC molecules are miscible with water. (2) A solid core of insoluble material (which may vary in size with the ambient humidity), surrounded by a single liquid phase 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.

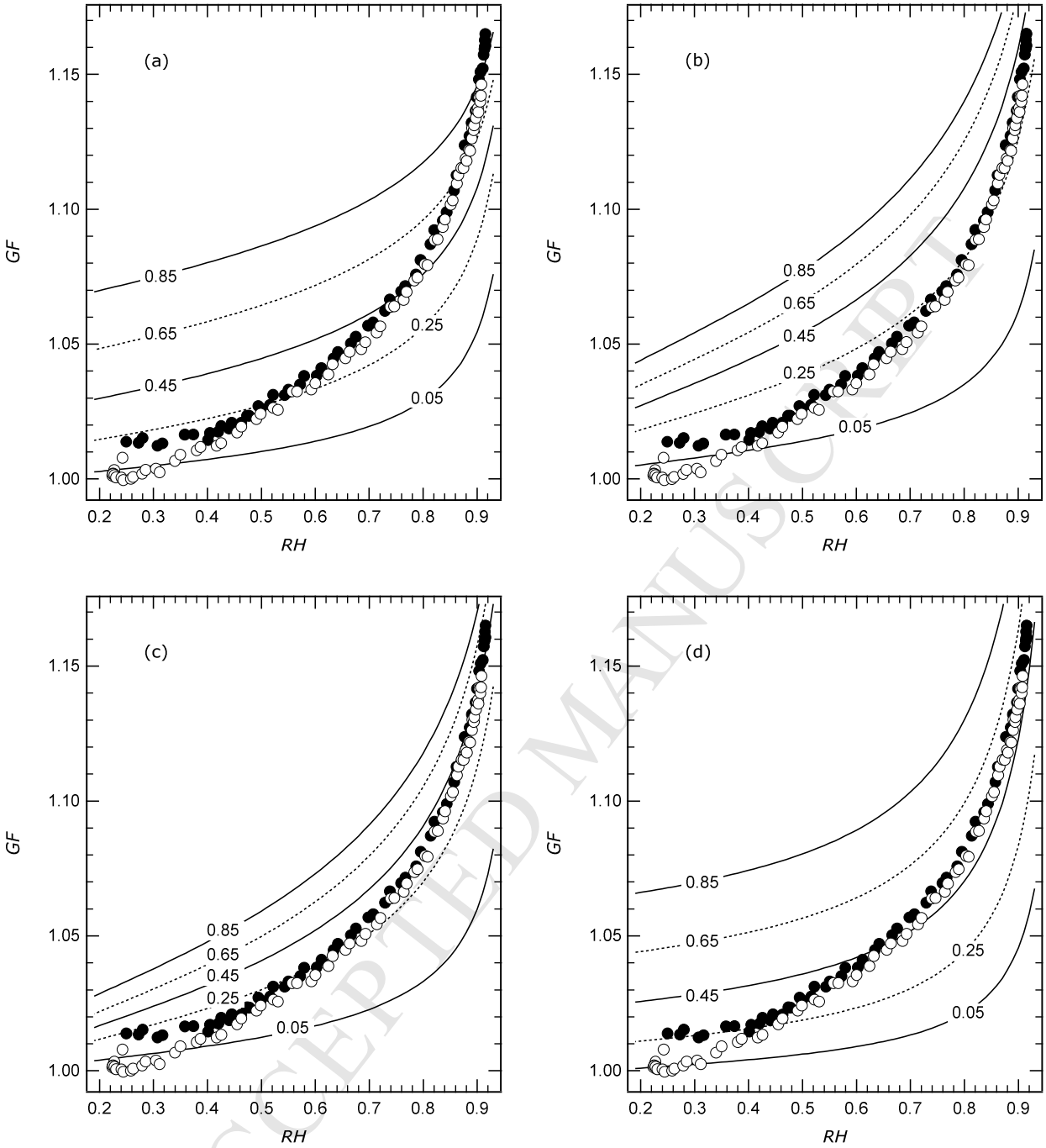



Figure 7. Growth factors (*GF*) of WSOC material in sample SX1, plotted against ambient relative humidity (*RH*), expressed as a fraction. Symbols: dots – deliquescence (low *RH* to high *RH*); open circles – efflorescence (high *RH* to low *RH*); lines - calculated values assuming different fixed mole 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).

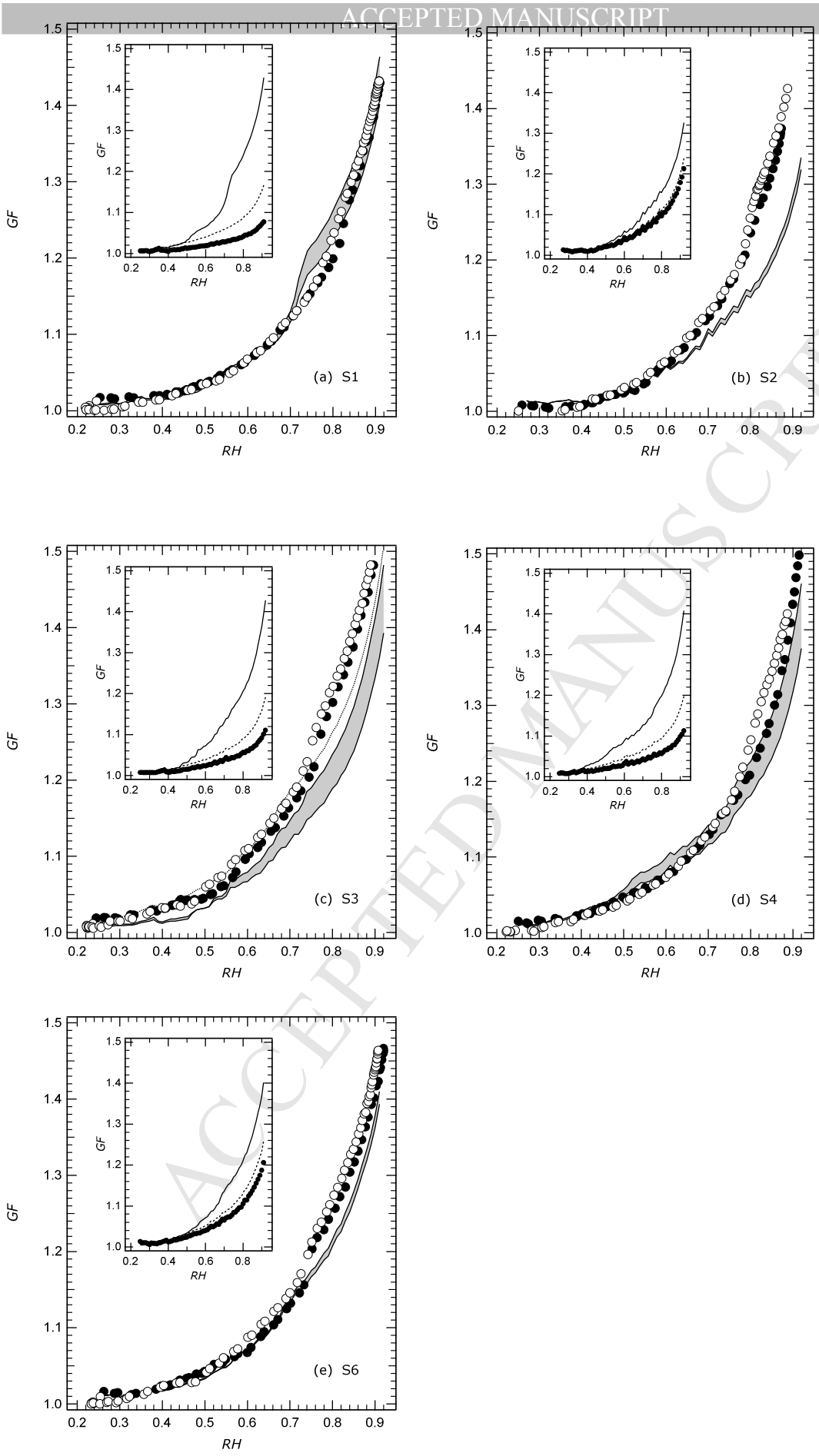



Figure 8. Measured and predicted growth factors (*GF*) as a function of relative humidity (*RH*). Symbols: dots – deliquescence measurement (scan from low *RH* to high *RH*); open circles – efflorescence

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).

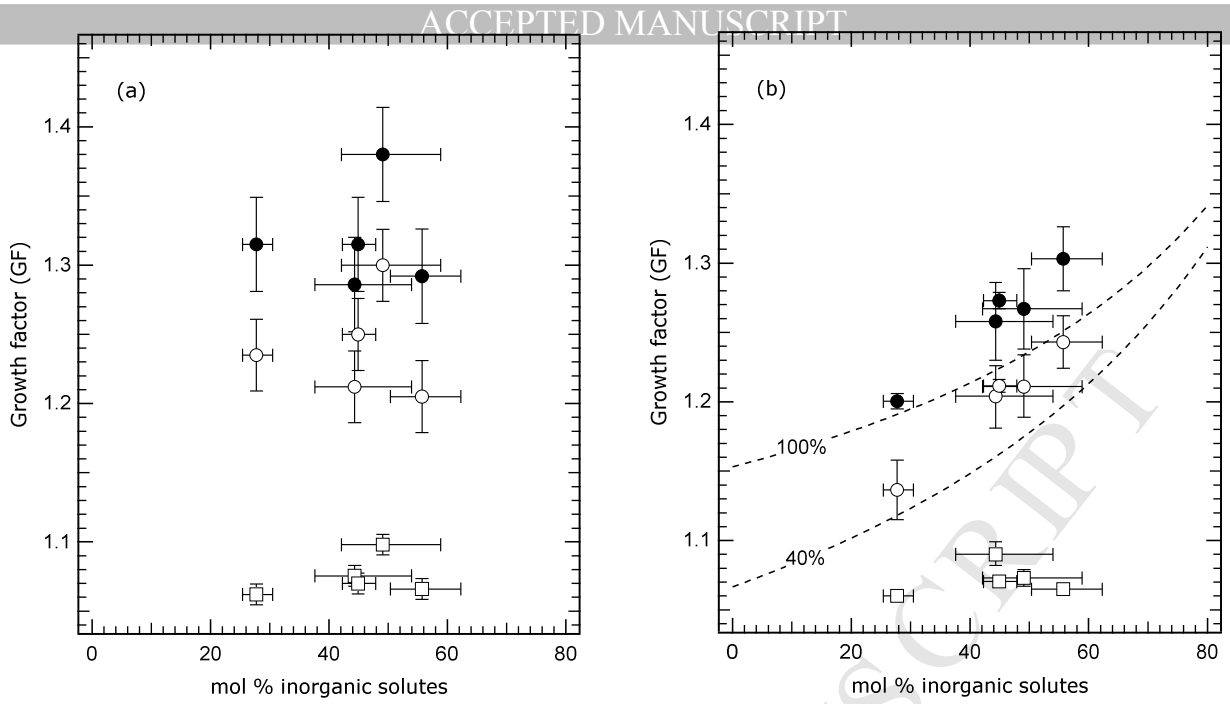



Figure 9. Growth factors (*GF*) for all samples except S5, at three relative humidities, plotted against the mol % of inorganic solutes present in the samples. (a) Measured values; (b) calculated (the same as in Figure 8). Symbols: open squares – 60% *RH*, open circles – 80% *RH*; dots – 85% *RH*. Lines: 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*.

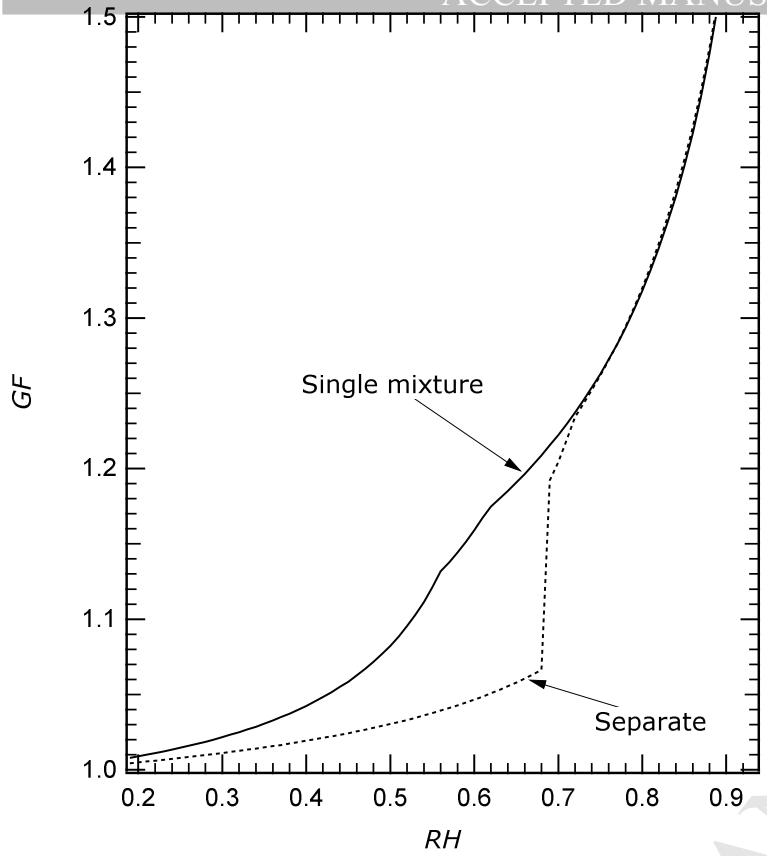



Figure 10. Calculated growth factors (*GF*) as a function of relative humidity (*RH*) for an aerosol consisting of 1 mole of a soluble organic compound (with water uptake conforming to Raoult's law), and 1 mole of $(NH_4)_{1.5}H_{0.5}SO_4$ (letovicite). Lines: solid – the mixture is modelled as a single 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*).

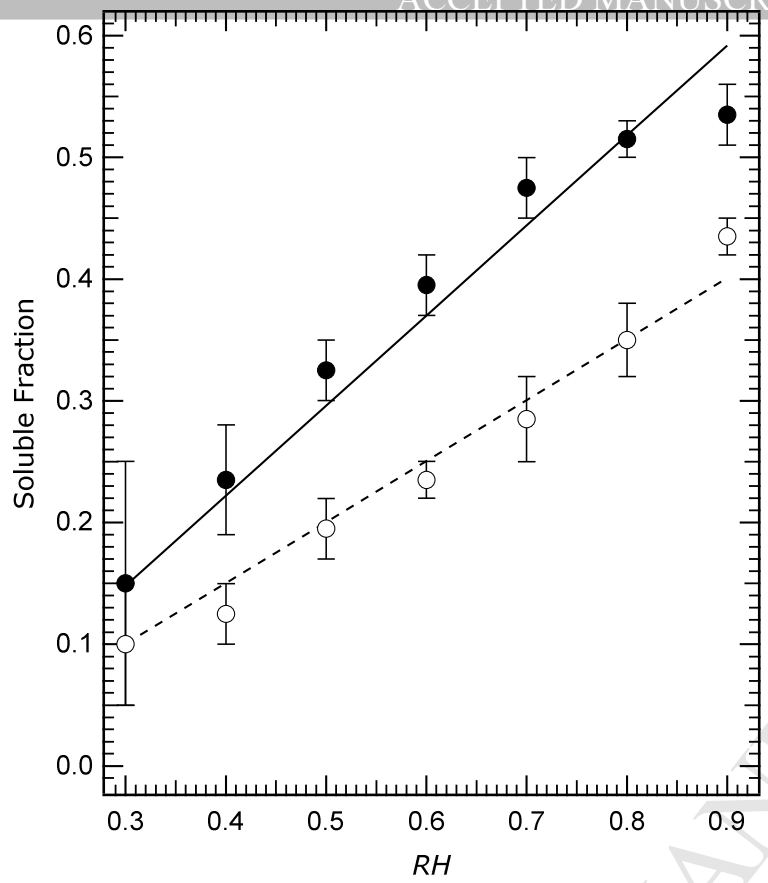



Figure 11. Soluble fractions of the WSOC material for which calculated growth factors agree with the measured values, plotted against relative humidity (*RH*). Symbols: dots – values for the assumption of Raoult's law mixing in the aqueous phase (Figure 7d); open circles – values from the UNIFAC calculation 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.