A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics

N. M. Donahue, S. A. Epstein, S. N. Pandis, and A. L. Robinson

Carnegie Mellon University, Center for Atmospheric Particle Studies, Pittsburgh, USA

Received: 23 September 2010 – Accepted: 25 September 2010 – Published: 15 October 2010

Correspondence to: N. M. Donahue (nmd@andrew.cmu.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

We develop the thermodynamic underpinnings of a two-dimensional volatility basis set (2-D-VBS) employing saturation concentration ($C^o$) and the oxygen content (O:C) to describe volatility, mixing thermodynamics, and chemical evolution of organic aerosol. This is an extension of our earlier one-dimensional approach employing $C^*$ only ($C^* = \gamma C^o$, where $\gamma$ is an activity coefficient). We apply a mean-field approximation for organic aerosol, describing interactions of carbon and oxygen groups in individual molecules (solutes) with carbon and oxygen groups in the organic-aerosol solvent. In so doing, we show that a linear structure activity relation (SAR) describing the single-component $C^o$ of a molecule is directly tied to ideal solution (Raoult’s Law) behavior. Conversely, non-ideal solution behavior (activity coefficients) and a slightly non-linear SAR emerge from off-diagonal (carbon-oxygen) interaction elements. From this foundation we can build a self-consistent description of OA mixing thermodynamics, including predicted saturation concentrations and activity coefficients (and phase separation) for various solutions from just four free parameters: the carbon number of a hydrocarbon with a $1\mu g m^{-3} C^o$, and the carbon-carbon, oxygen-oxygen, and non-ideal carbon-oxygen terms. This treatment establishes the mean molecular formula for organics within this 2-D space as well as activity coefficients for molecules within this space interacting with any bulk OA phase described by an average O:C.

1 Introduction

It is well established that atmospheric organic aerosol (OA) exists as a complex mixture of thousands of individual organic compounds (Hildemann et al., 1991; Fraser et al., 1997, 1998; Goldstein et al., 2008). Furthermore, it is clear that a large fraction of OA consists of organic compounds that are difficult to separate in a gas chromatograph (Schauer et al., 1996), and that very often a majority of OA consists of quite highly oxidized organic molecules (Zhang et al., 2007; Aiken et al., 2008; Ng et al., 2009).
As an example, the most abundant identified organic molecule during the Pittsburgh Air Quality Study, generally a carboxylic acid, typically comprised about 1% of the total organic carbon mass (Robinson et al., 2006; Shrivastava et al., 2007). A typical 200 nm diameter particle contains 5–10 million individual organic molecules with molar weights around 200 g mole\(^{-1}\). The most abundant molecule appears approximately 50,000 times in that single particle, while hundreds to thousands of molecules appear of order 100–1000 times each in that same particle.

A critical determinant of whether an organic compound is found in OA (as opposed to the gas phase) must be saturation vapor pressure. Here we shall use saturation mass concentration \((C^0 \text{ and } C^*, \text{ in } \mu g m^{-3})\) because mass measurements pervade aerosol science, and we shall use the terms “volatility” and “saturation concentration” interchangeably. The difference between \(C^0\) and \(C^*\) is that \(C^0\) is the pure (sub-cooled) liquid saturation compound of a substance, while \(C^*\) includes the activity coefficient in a mixture \(γ\), thus \(C^* = γC^0\). Furthermore, because volatility spans an enormous range, we shall generally use a log axis, employing the shorthand \(lC^0 = \log_{10}C^0\) and \(lC^* = \log_{10}C^*\).

We have previously presented a framework describing OA mass – whether or not the constituents are identified – with a series of volatility “bins” that are separated by one decade \((δ/lC^* = 1)\), which we shall refer to as the “one-dimensional volatility basis set” (1-D-VBS) (Donahue et al., 2006). Here we shall extend the 1-D-VBS to a second dimension in order to improve our ability to predict the thermodynamics, including organic mixing and hydrophilicity, and ultimately to coherently describe the oxidation chemistry.

Because OA is a complex mixture, the thermodynamics of mixtures are important (Pankow, 1994). From both a modeling and a theoretical perspective, it is desirable to be able to predict volatility under ambient conditions. It is also important to know whether different organics will tend to form a single, homogeneous phase or whether they will phase separate. One reason is that phase separation can sharply reduce the predicted mass of the less abundant phase (Bowman and Melton, 2004), while...
another is that separate phases will allow distinct OA types to remain externally mixed (for example with different sized modes) while a single OA phase will tend to drive OA composition toward a uniform, internal mixture (Asa-Awuku et al., 2009).

The challenge we seek to confront here is that the molecular composition of OA is unknown and very complex. Consequently, it will be useful to develop a thermodynamic treatment based on OA properties that we can directly observe and constrain, to complement approaches based on surrogate molecules. Each method contains intrinsic errors – the simplified thermodynamics we shall present here cannot represent the free-energy change of individual molecules upon solvation, but the individual molecules used in surrogate approaches may or may not adequately represent those of actual OA mixtures.

With the surrogate approach, one strives to carefully select a small set of surrogate molecules and then to calculate the thermodynamics as accurately as possible. The universal functional activity coefficient (UNIFAC) method (Fredenslund et al., 1975) has been widely used to describe the activity of organic aerosol mixtures (Koo et al., 2003; Bowman and Melton, 2004; Pun et al., 2006). UNIFAC is a group contribution method that gives the excess Gibbs energy of each component in a specified mixture based on the non-ideality arising from the size and shapes of the constituent molecules and the non-ideality arising from their energetic interactions (Asher et al., 2002; Chandramouli et al., 2003). UNIFAC calculations require knowledge of the aerosol mixture composition and the abundance of each of the molecular subgroups (Bowman and Melton, 2004).

UNIFAC calculations of some surrogate species also present challenges. The multifunctional and long-chain molecules, which are typically found in the atmosphere, are not accurately described by the UNIFAC method (Koo et al., 2003; Bowman and Melton, 2004). Also, UNIFAC calculations can be computationally intensive (Bowman and Melton, 2004). Finally, it is not obvious that surrogate molecules will accurately represent complex ambient OA mixtures. For instance, Clegg et al. (2008a,b) have shown that very similar seeming organics can have very different activities in model mixtures,
and thus that it is difficult, a priori, to choose “good” surrogates for OA mixtures.

The approach developed in this paper relies principally on data from in situ mass spectrometers that reliably measure a large fraction of the OA mass and are able to constrain the bulk atomic composition (C, H, O and sometimes N, S,...) with good precision (Aiken et al., 2008). The most precisely constrained measure is the oxygen to carbon ratio (O:C). Here we shall show that we can develop a self-consistent thermodynamic framework for OA mixtures based on the total OA mass \( C_{OA} \) and O:C. Our key hypothesis will be that we can formulate the thermodynamics based on a mean-field approximation in which individual molecules interact with aggregate, average (bulk) properties of the OA solvent. This is motivated by the example of composition presented above – individual molecules in an OA particle are rarely if ever solvated by themselves but rather by a complex mixture of many molecules containing several different functional groups in many combinations. While the individual solvent molecules can not be completely enumerated, the necessary average properties can now be measured routinely.

Both a mean structure activity relation (SAR) connecting measured \( C^o \) and O:C with composition (i.e. \( n_C, n_H, \) and \( n_O \) for that \( C^o \) and O:C), and a generalized prediction of activity coefficients for individual organics interacting with a homogeneous OA mixture will fall out naturally and self-consistently from this formulation. This rich behavior emerges from only four parameters: a compound with a reference volatility (we use pentacosane), and three interaction terms expressing in turn carbon-carbon interactions (non-polar interactions), oxygen-oxygen interactions (polar interactions), and the non-ideality of the carbon-oxygen cross interaction (the hydrophobicity). We shall explore the implications of this, both in the laboratory, where different organics may or may not serve as good “seeds” for other organics (Song et al., 2007; Asa-Awuku et al., 2009; Vaden et al., 2010), and in the atmosphere.

It is very important to understand our objectives. We do not seek to accurately predict the phase partitioning of specific molecules (e.g. levoglucosan, etc.). Rather, we seek to understand the bulk behavior of OA. We will use this thermodynamic treatment within
the VBS, where the phase partitioning itself is known almost by definition. We do want to be able to invert the problem – to find \( n_C, n_O, n_H \), etc. for material with a certain \( C^o \) and O:C. We also want to be able to understand the physical properties of the OA, including hygroscopicity as well as the potential for different classes of OA to phase separate, and we want to inform chemical mechanisms describing how these properties (including volatility and O:C) evolve during photochemical oxidation.

## 2 Theory

### 2.1 Saturation concentrations and vaporization enthalpies

Volatility is controlled principally by the vaporization enthalpy of a substance, especially over liquids and other disordered condensed phases where the vaporization entropy is nearly constant (Prausnitz et al., 1998). We recently showed that this holds for a wide range of organics. Specifically, a pseudo-Arrhenius expression can describe the temperature dependence of \( C^o \), and there is a nearly linear relationship between \( lC_{300}^o \) (\( \log_{10} C^o \) at 300 K) and the vaporization enthalpy (Epstein et al., 2010). Here we shall assume that the relationship is precise.

Because we use Boltzmann terms extensively, we shall scale energies by the gas constant and report them in Kelvins, \( \theta = E/R \). Thus the vaporization enthalpy becomes \( \Delta g^{vap} \). A key value is the change in \( \theta \) driving a one decade change in \( C^o \) at 300 K: \( \Delta \theta_{10} = d \Delta \theta^{vap}/d lC^o \). We also need to know the vaporization enthalpy for \( C^o = 1\mu gm^{-3} \): \( \Delta \theta^{vap}_{1\mu g} \). So, for a compound \( i \):

\[
\Delta \theta^{vap}_i = \Delta \theta^{vap}_{1\mu g} - \Delta \theta_{10} \times lC_{i,300}^o \quad (1)
\]

or

\[
lC_{i,300}^o = \left( \Delta \theta^{vap}_{1\mu g} - \Delta \theta^{vap}_i \right) \frac{1}{\Delta \theta_{10}} \quad (2)
\]
This means that we need to determine or specify only one of $\Delta \theta_{i}^{\text{vap}}$ or $I C_{i,300}^{\theta}$ and we know the other.

Simple theoretical considerations suggest $\Delta \theta_{10} \simeq -690 \text{K} \left( -5.7 \text{kJ mole}^{-1} \right)$ (Donahue et al., 2006): $\Delta \theta_{10}$ should cause a one-decade change in a Boltzmann term at 300 K. On the other hand, empirical correlations give a larger value of $\Delta \theta_{10} \simeq 1500 \text{K} \left( -13 \text{kJ mole}^{-1} \right)$ (Epstein et al., 2010). Whatever the correct value of $\Delta \theta_{10}$, an important empirical finding is that it depends almost exclusively on $I C_{10}^{\theta}$ and it is independent of the degree of substitution (oxidation) of an organic molecule (Epstein et al., 2010). Consequently, we shall make the assumption that the vaporization enthalpy is simply the energy required to remove a given organic molecule from an organic solvent.

### 2.2 Describing vaporization energies

It is useful to start with a very basic formulation: far more detailed treatments can be found in any advanced thermodynamics text (e.g. Prausnitz et al., 1998). The steps involved in evaporating from a solvent are shown in Fig. 1. Specifically, the overall energy change includes the energy required to remove a molecule $i$ from a solvent $s$ ($\theta_{i,s}$), but also the change in energy in the solvent when the interaction energy of the molecule with the solvent is replaced by the interaction energy of the solvent with itself ($\theta_{s,s}$). What this boils down to (so to speak), is

$$\theta_{i,s}^{\text{vap}} = \theta_{i,s} + (\theta_{i,s} - \theta_{s,s}) = 2 \theta_{i,s} - \theta_{s,s} \quad (3)$$

Note that when the molecule is the solvent (a single component system), the term in parentheses vanishes and we are left with the simple and intuitive result of $\theta_{i}^{\text{vap}} = \theta_{i,i}$. This defines $C^{\theta}$ via Eq. (2).

We can also use Eq. (2) to define $C^{\ast}$ when $i \neq s$, meaning the excess entropy of $i$ is approximately zero (Prausnitz et al., 1998). For an ideal solution, $\gamma = 1$ and $C^{\ast} = C^{\theta}$, therefore $\Delta \theta_{i,s}^{\text{vap}} = \theta_{i,i}$. What we really want to know is the difference between the
vaporization energy of \( i \) in the solvent and the vaporization energy of pure \( i \):
\[
\delta \theta_{i,s} = 2\theta_{i,s} - \theta_{s,s} - \theta_{i,i} \tag{4}
\]
For an ideal solution \( \delta \theta_{i,s} = 0 \), and we obtain:
\[
\theta_{i,s}^{\text{ideal}} = \frac{1}{2}(\theta_{s,s} + \theta_{i,i}) = \bar{\theta} \tag{5}
\]
This is a very general result: for ideal solutions, the cross term is the average of the corresponding diagonal terms. We will therefore define this cross term in terms of the deviation from the mean:
\[
\theta_{i,s} = \bar{\theta} + \delta \theta_{i,s} \tag{6}
\]
which is part of an interaction matrix of this form:
\[
\theta = \begin{pmatrix} \theta_{i,i} & \bar{\theta} + \delta \theta_{i,s} \\ \bar{\theta} + \delta \theta_{i,s} & \theta_{s,s} \end{pmatrix} \tag{7}
\]
All of this can be combined to obtain the energy of \( i \) in \( s \):
\[
\theta_{i,s}^{\text{vap}} = \theta_{i,i} + 2\delta \theta_{i,s} \tag{8}
\]
We have already shown the exponential relationship between \( \theta \) and both \( C^0 \) and \( C^* \) in Eq. (2), so at an equivalent level of approximation the excess enthalpy (Prausnitz et al., 1998) is just \( 2\delta \theta_{i,s} \), and the activity coefficient of \( i \) in \( s \) is given by a Boltzmann expression:
\[
\gamma_{i,s} = \exp\left(-\frac{2\delta \theta_{i,s}}{T}\right) \tag{9}
\]
This is, quite simply, why activity coefficients are so difficult to predict or even parameterize – an energy difference of only \( kT \), (about 2.5 kJ mole\(^{-1}\) at 300 K), is enough to
make the activity coefficient a factor of \( e \), and because \( \gamma \) is exponential it will grow (or shrink) very rapidly with increasing \( |\delta \theta| \). Treatments attempting to predict activity coefficients for individual molecules must carefully account for multiple thermodynamic terms (Jang et al., 1997; Pankow and Chang, 2008); that is appropriate for those specific situations, but the treatments require detailed knowledge of the solvent composition as well as the exact molecule of interest. We have neither.

### 2.3 Simple mixtures

So far the solvent has been of unspecified composition. Now let us consider a molecule \( A \) \((i \rightarrow A)\) in a solution \( s \) that is a two-component mixture of molecules \( A \) and \( B \), where the fraction of \( A \) in the solution is \( f^s_A \) and the fraction of \( B \) is \( f^s_B = (1 - f^s_A) \). This means that, from the perspective of a molecule \( A \), it will be interacting with a “solvent” consisting of an \( A-B \) mixture. If the mixture is homogeneous and isotropic, meaning that the local fractions (and orientations) of the molecules are the same throughout the mixture, then the interaction energy an individual molecule \( A \) with the solvent is

\[
\theta_{i,s} \rightarrow \theta_{A,s} = f^s_A \theta_{A,A} + (1 - f^s_A) \theta_{A,B}
\]  

(10)

From Fig. 1, we also need to know the interaction energy of the solvent with itself. This is a bit more complicated. We shall assume that the solvent “molecule” replacing \( A \) when it evaporates is a composite containing \( f^s_A \) of molecule \( A \) and \( (1 - f^s_A) \) of molecule \( B \), each of which will interact with the fractions of \( A \) and \( B \) in the solvent.

\[
\theta_{s,s} = f^s_A \theta_{A,A} + 2 f^s_A (1 - f^s_A) \theta_{A,B} + (1 - f^s_A)^2 \theta_{B,B}
\]  

(11)

We are now in a position to calculate \( \delta \theta_{A,s} \), the excess energy of \( A \) in \( s \). In a recurring theme, it can be shown that for an ideal mixture, when \( \theta_{A,B} = 1/2 (\theta_{A,A} + \theta_{B,B}) = \bar{\theta} \), then \( \delta \theta_{A,s} = 0 \). Consequently, for non-ideal mixtures with \( \theta_{A,B} = \bar{\theta} + \delta \theta_{A,B} \):

\[
\delta \theta_{A,s} = (1 - f^s_A)^2 \delta \theta_{A,B}
\]  

(12)
In the end, the activity coefficient of $A$ in an A-B mixture is just what one might expect – it depends on the non-ideality of the off-diagonal interaction element, exactly as in Eq. (9). The total excess energy is now a function of the solution composition, increasing quadratically as the solvent becomes increasingly dissimilar to $A$.

In defining these mixtures as homogeneous and isotropic, we have made our second major assumption. Real molecules (for instance water) don’t interact randomly with other molecules. However, given that there are few if any observables for OA to contradict this assumption, it is either robust or just necessary. The very complexity of ambient OA may make it more homogeneous and isotropic than simpler mixtures. The extreme example is the tendency of mixtures to form liquid or amorphous liquid-like mixtures where the individual constituents would crystalize (Cappa et al., 2008; Zobrist et al., 2008), but much more generally the complex mixtures found in atmospheric OA will smooth out the rough edges of highly specific A-B interactions.

### 2.3.1 Non-ideal activity

It is worth exploring the activity coefficient a bit further. What we are really interested in is the activity of a compound, $\alpha_i$, defined relative to a pure reference state in vapor-liquid phase equilibrium and related to the activity coefficient and the fraction of $A$ in the solution, $f_A$. This is simply:

$$\alpha_A = f_A \gamma_{A,s} = f_A \exp\left(-2\delta\theta_{A,s}/T\right)$$  \hspace{1cm} (13)

for the two-component mixture this becomes

$$\alpha_A = f_A \exp\left(-2 \left(1 - f_s^A\right)^2 \delta\theta_{A,B}/T\right)$$  \hspace{1cm} (14)

The consequences of this are shown in Fig. 2. The bottom line is that non-ideal behavior emerges very quickly for quite modest values of $\delta\theta \approx T$, and that this will induce phase separation. However, the real systems we are interested in consist of very complex mixtures. Therefore, we must ask whether it is possible to predict these small deviations from ideality with anything close to sufficient accuracy.
An ansatz here is the complex mixtures found in atmospheric OA will be sensitive to mean-field properties, and that we can indeed describe this mean-field behavior usefully. For any individual molecule the values we predict will not be accurate, but we hypothesize that on average they will be (with some semi-empirical calibration). However, it is very important to appreciate the contrapositive of this claim: predictions based on a few surrogate molecules are almost certain to be inaccurate for complex mixtures, because the unique (and therefore unrepresentative) properties of those individual surrogate molecules will likely not represent the average (bulk) properties of the mixture very well.

3 Structure activity relations

3.1 Data and mean trends

With the theoretical foundations of our approach developed, we shall now make what may seem like an abrupt turn. We would like to be able to describe $IC^*$ in terms of observable average molecular properties of OA – specifically the average carbon, hydrogen, and oxygen numbers ($n_C, n_H, n_O$). Initially we will focus on $IC^0$ – the pure saturation concentrations. There are a number of structure activity relations in the literature; one that is especially convenient in this context is SIMPOL (Pankow and Asher, 2008), in which (log$_{10}$ of) the sub-cooled liquid vapor pressure of a pure substance can be described as a simple sum of contributions from various functional groups:

$$\log_{10}(p_{L,i}^0) = b_0 + \sum_k v_{k,i} b_k$$

(15)

where $b_k$ is the effect of the $k$–th “structural element” and $v_{k,i}$ is the number of those elements in the molecule $i$. That is tantalizingly close to what we want. Here we shall build a SAR akin to SIMPOL with the minor difference of $IC^0$ for log$_{10}p_{L,i}^0$, focusing on the carbon and oxygen numbers ($n_C$ and $n_O$) of the organics. Specifically, our objective...
in this paper is to describe $1C^O = f(n_C, n_O)$, neglecting for the time being other contributions (for example, N, S) in favor of the “bedrock” oxygenated organic compounds that appear to comprise the bulk of OA.

Data for a very wide range of compounds are plotted vs. $n_C$ in Fig. 3. These form the basis for our SAR. The data are from NIST as well as the literature (Koponen et al., 2007; Cappa et al., 2007). The oxygenated compounds all have functionality identified in ambient OA (-OH, =O, -C(O)OH, etc.). We group molecules in classes defined by functionality (i.e. mono-alcohols, with a single -OH), and plot each class with a different symbol. The broad, colored bands in the background are volatility classes we use elsewhere (Donahue et al., 2009); here they serve to guide the eye through the various figures. The important attribute for atmospheric partitioning is that compounds in the light-green range will be semi-volatile in the atmosphere, existing in significant fractions in both the condensed and vapor phases. Compounds in the light red and gray ranges will be almost completely condensed under typical ambient conditions. Note that there are relatively few data in the semi-volatile (green) range and almost none below that. This is because a saturation concentration of 1µg m$^{-3}$ corresponds to approximately 10$^{-7}$ torr; this is an exceedingly low value that is difficult to measure.

There are several shortcomings with the available data. The data in Fig. 3 become noisy approaching the semi-volatile range and they do not include richly multifunctional compounds such as those that must comprise a large portion of the OOA pool. Some of the noise at a low $C^O$ may be related to uncertainty in converting to sub-cooled liquid values, and the rest is likely due to measurement uncertainties associated with very low vapor pressures. The diacid series does reach these low values, partly because the atmospheric community has devoted considerable effort to measuring their vapor pressures (Tao and McMurry, 1989; Bilde and Pandis, 2001; Bilde et al., 2003; Koponen et al., 2007; Cappa et al., 2007, 2008).

The most obvious feature of Fig. 3 is the succession of parallel lines formed by the different classes. This means that there is a well-defined effect of increasing $n_C$, independent of molecule class. We can thus define a robust relationship between $n_C$ and
$C^o$: each class of molecule has an $IC^o$ vs. $n_C$ slope of 0.475 decades per carbon. This defines the carbon-number dependence of $IC^o$ for our SAR, which is the slope of the trend lines in Fig. 3. We also need a reference point, which we will define as the carbon number $n_C$ at $C^o = 1 \mu g m^{-3}$ for pure hydrocarbons. From Fig. 3 this is pentacosane (C$_{25}$), indicated with the black square. This reference $n_C$ may be uncertain by 2 carbons (corresponding to a 10% uncertainty in the group contribution for carbon to $C^o$). Some data for high-$n_C$ alkanes lie above the trend line, while others fall on it. Given the precision of the relationship for lower-$n_C$ alkanes, we believe this indicates the difficulty in measuring such low vapor pressures, and that the lower envelope represented by our SAR represents the true values for the $n$-alkanes.

The next step in developing the SAR is to define the influence of oxygenation. A challenge is that we need to account for the influence of different functionalities. Increasing functionalization decreases volatility systematically, with logarithmic factors given by the vertical offsets of the various trend lines in Fig. 3 from the pure hydrocarbon trend line. For example, each carbonyl group (=O, blue and yellow symbols) decreases $C^o$ by 1 decade, and each alcohol group (-OH, green symbols) decreases $C^o$ by 2.3 decades. One organic-acid group decreases $C^o$ by about 3.5 decades, or 1.75 decades per oxygen (the carbon is not counted because it is included in $n_C$), which is very close to a simple sum of (-OH) and (=O).

These effects are shown in a different way in Fig. 4, which casts the effects of functionalization in a 2-D space with $C^o$ on the x-axis and O:C on the y. The different functional groups generate different slopes (shown with dashed red lines). The wedges in Fig. 4 describe the range of values one might expect as a carbon backbone is progressively functionalized. These wedges thus represent the effect of oxidation chemistry on the volatility and O:C of a reduced precursor, under the limiting condition that oxidation only functionalizes the backbone (i.e. $n_C$ remains unchanged).

Despite the uncertainties in the underlying data, there are enough data to suggest that these single-substituent SAR effects on $C^o$, described by the trend lines in Figs. 3 and 4, are not grossly erroneous. In the VBS we group compounds in bins separated by
an order of magnitude in $C^o$, and overall we estimate the uncertainty in extrapolation in these simple SAR calculations to be somewhere between a factor of 10 and 100, meaning ±1 bin to either side of the estimated volatility bin.

If we are to construct a two-dimensional thermodynamic framework for bulk OA, we need to describe an average effect of added oxygen. Multiple lines of evidence suggest that (-OH), (=O) and (-C(O)OH) functionalities are major contributors to ambient oxygenated OA, known as OOA (Zhang et al., 2007). First, OOA is associated with large signals in the AMS at $m/z=18$ ($H_2O^+$) and 44 ($CO_2^+$), consistent with dehydration and decarboxylation of organic acids (Aiken et al., 2008). Second, NMR analyses of solvent-extracted filter samples suggest high abundances of these functionalities (McFiggans et al., 2005; Moretti et al., 2008). Third, FTIR analyses of ambient filter samples show significant contributions of (-OH), (=O) and (-C(O)OH) functional groups (Hawkins et al., 2010). Finally, a van Krevlen analysis of ambient OA by Heald et al. (2010) is consistent with either hydroxycarbonyl or organic-acid functionalization dominating organic-aerosol oxidation on average (i.e. an average of one hydrogen is lost for every oxygen added to ambient OA).

Based on this evidence, we shall assume that, on average, OA consists of multifunctional organics containing organic-acid terminal groups and an equal mixture of alcohols and ketones along the backbone. We can thus define the average effect of each added oxygen as decreasing $C^o$ by 1.7 decades, shown in Fig. 3 by thick, dashed lines and in Fig. 4 with black lines. This is consistent with either acid or hydroxycarbonyl functionality, provided that these simple SAR values can be extrapolated to multifunctional compounds that presumably constitute OOA. The end result of this analysis is a simple, three-parameter structure activity relation:

$$lC^o = \left(n_0^C - n_C\right) b_C - n_O b_O$$

(16)

with $n_0^C = 25$, $b_C = 0.475$ and $b_O = 1.7$.

Other functionalities are not used to establish the SAR, but they fall within the range already defined. Hydroperoxides (-OOH) are assumed to be important under low-NO$_x$
conditions (Logan et al., 1981; Donahue and Prinn, 1993), and we see evidence for their formation in 2-D-NMR samples of terpene ozonolysis (Maksymyiuk et al., 2009). The effect of -OOH on C^0 is similar to -OH (Pankow and Asher, 2008).

Ethers (-O-) may also be significant in ambient OA. They have a modest effect on volatility, slightly less than the =O effect, of about 0.7 orders of magnitude per substituent (Pankow and Asher, 2008). Ethers may form through ring-closure reactions during the oxidation of larger organic compounds (Lim and Ziemann, 2005) or as part of accretion reactions involving small monomers such as glyoxal (Kroll et al., 2005; Carlton et al., 2006; Volkamer et al., 2007; Galloway et al., 2009). Again, as their effect is not far outside of the =O line (even counterbalancing the -OOH effect), we do not believe that ethers will ruin the relation we propose.

The most significant omission in Fig. 3 is the volatility of organic nitrates. Organic nitrates are certainly important in high-NOx environments (Lim and Ziemann, 2005; Presto and Donahue, 2006; Zhang et al., 2006). A nitrate group (-ONO_2) typically reduces vapor pressure by about 2.5 orders of magnitude (Pankow and Asher, 2008), so at a similar level of approximation we could introduce b_N = 2.5 to a more complete SAR. In any case, nitrates and any other special functionality (organo sulfates, amines, alkenes, etc.) will need to be treated separately if they are to be modeled.

The diacids reveal the major shortcoming of this simple, linear SAR. While there is significant uncertainty in their volatility, the diacids (solid magenta) consistently fall about 1 order of magnitude below the 4-oxygen (dashed green) trend line of our simple SAR, which should be very close to the value for two acid groups. This discrepancy will later provide an important constraint as we relax the ideal, linear assumptions we have used so far.

3.2 Defining the 2-D volatility space

The SAR described using Eq. (16) allows us to determine the mean properties (specifically n_C) in the 2-D space presented in Fig. 4. The mean-property lines of 1.7 decades per oxygen are shown in black. The C_{20} lines are much shallower, as O:C increases...
much more slowly with each added O for larger molecules. It is very important to re-
alize that we are not using this SAR to predict $IC^O$ for a given $n_C$ and O:C but rather
the reverse. In both laboratory experiments and field measurements we typically have
a much better constraint on O:C and $IC^O$ than $n_C$, and so we are using the SAR to
infer $n_C$ (and the average molecular formula) from those observables. Likewise, in the
VBS modeling framework (Donahue et al., 2006), we use $IC^*$ explicitly and are thus
not directly sensitive to uncertainties in volatility.

This SAR allows us to construct a family of contours (isopleths) for constant values
of $n_C$, $n_H$, and $n_O$, with some uncertainty and spread in the $IC^O$, O:C space. This is
shown in Fig. 5. These contours interconnect predictive, diagnostic, and prognostic
aspects of the 2-D space. This paper addresses the predictive aspect, focusing on the
relationships between $IC^O$ and O:C on the one hand and $n_C$, $n_H$, and $n_O$ on the other.
In subsequent publications we will explore the diagnostic and prognostic aspects. The
diagnostic aspect allows us to evaluate the evolution of $n_C$, $n_H$, and $n_O$ in observational
(laboratory and field) datasets where $IC^*$, O:C, and total organic mass are constrained,
using the relationships established here. The prognostic aspect allows us to predict
the evolution of $IC^*$, O:C, and total organic mass in box and transport models, with
individual reaction steps in part constrained by the SAR developed here.

Figure 5 shows mean $n_C$ as diagonal black isopleths, mean $n_O$ as green curving
isopleths, and the sum of oxygen and carbon $n_M = n_C + n_O$ as cyan curving isopleths
(this will be important later, as $n_M$ is similar to the molar volume). The carbon isopleths
are labeled, and for O:C = 1:1 they intersect at $n_C = n_O = 0.5n_M$ (labeled in bold at 2,
4, and 6). These isopleths define the average composition in this 2-D space. A notable
feature is that the broad region of our 2-D space with $10^{-5} < C^O < 100\mu g m^{-3}$ and $1:3
< O:C < 1:1$ (more or less the OOA range) comprises compounds with 5–12 carbons
and 4–8 oxygens. Even though this is also the most uncertain region, based almost
completely on extrapolation, the range in carbon and oxygen numbers in this region is
not so very large.
3.3 Solvation energies in the SAR

Though the ideal-solution and structure-activity treatments presented so far may seem disconnected, they are not. Our next objective is to relate the SAR to interaction energies of the functional groups in a molecule. We will start with the linear SAR we just defined in Eq. (16), seeking to define the specific interaction energies between functional groups in a molecule with the functional groups in the solvent (the bulk OA). We shall demonstrate that the linearity of the SAR and ideal solution behavior given by $\delta \theta_{A,B} = 0$ in Eq. (12) are tightly linked. Using this information, we shall introduce a single non-ideal interaction energy that simultaneously makes the SAR slightly non-ideal (correcting the overestimation of diacid volatility) and the solution non-ideal, introducing realistic hydrophobic interactions leading to phase separation of non-polar and polar organics in some cases.

In keeping with the simple volatility SAR, our central assumption is that we can calculate the solvation energy of an organic constituent as a linear combination of interaction energies between different functional groups in the molecule and different functional groups in a solvent, and that in a rich, amorphous system, these interactions will be homogeneous and isotropic. In other words we will model OA as an amorphous collection of functional groups, defined by the mean properties of the OA and without any other regard to the molecular composition of the OA. The central assumption of this work is that we can use a mean-field approximation to treat the interaction of individual organic molecules with the complex organic solvent otherwise known as organic aerosol.

In this paper, we have made the simplifying assumption that we can reduce low-NOx OA to two “functional groups”: carbon and oxygen. The average composition of the OA solvent is given by the carbon and oxygen fractions, $f_C^s$ and $f_O^s = (1 - f_C^s)$. We thus need only know two pairwise interaction energies $\theta_{CC}$, $\theta_{OO}$, and the cross term $\theta_{CO}$. The situation is very much the same as with the simple solvation model presented in Eqs. (10) and (11), but now the solvated molecule $i$ is in essence a polymer string of
different functionalities describable by \( f^i_C \) and \( f^i_O = \left( 1 - f^i_C \right) \), and the solvent is a mixture of such molecules with a wide range of individual \( f^i_C \) and \( f^i_O \) values but a well-defined average given above. We assume that it is the functionalities of the solvent and not the specific molecules of the solvent that dominate the solute-solvent interactions.

The situation is depicted in Fig. 6. Applying this to organic aerosol, for a given OA component \( i \) in the bulk OA solvent \( s \):

\[
\theta_{i,s} = n^i_C \left( f^i_C \theta_{CC} + f^i_O \theta_{CO} \right) + n^i_O \left( f^i_C \theta_{CO} + f^i_O \theta_{OO} \right) \\
= n^i_M \left[ f^i_C \left( f^s_C \theta_{CC} + (1 - f^s_C) \theta_{CO} \right) + (1 - f^i_C) \left( f^s_C \theta_{CO} + (1 - f^s_C) \theta_{OO} \right) \right]
\]

\( 17 \)

where we have used \( n^i_C = n^i_M \cdot f^i_C \) and \( n^i_O = n^i_M \cdot f^i_O \), the fractions of carbon and oxygen in the molecule and solvent.

When the component \( i \) is dissolved in itself we have a single-component system defined by:

\[
\theta_{i,i} = n^i_C \left( f^i_C \theta_{CC} + f^i_O \theta_{CO} \right) + n^i_O \left( f^i_C \theta_{CO} + f^i_O \theta_{OO} \right) \\
= n^i_C f^i_C \theta_{CC} + 2 \frac{n^i_C n^i_O}{n^i_C + n^i_O} \theta_{CO} + n^i_O (1 - f^i_C) \theta_{OO}
\]

\( 18 \)

However, there are two key assumptions in any linear SAR, including Eq. (16). First, functional-group contributions are assumed to be independent of the specific molecular structure (the location and proximity of the functional groups in a molecule), and, second, the contribution of a functional group to volatility is due only to its own abundance (and not to the abundance of other functional groups). In other words, addition of an (-OH) functionality to an organic molecule will suppress the volatility by 2.3 decades, regardless of the rest of the molecular structure. As we are assuming that volatilities
and solvation energies have a one-to-one relationship, the linear SAR is equivalent to:

$$\theta_{i,i} = n_C \theta_{C,C} + n_O \theta_{O,O}$$  \quad (19)$$

with no apparent cross term, $\theta_{C,O}$ (as that would violate the second assumption just stated). Equations (18) and (19) can be equated, which yields

$$\theta_{\text{linear}}^{\text{CO}} = \frac{1}{2} (\theta_{CC} + \theta_{OO}) = \bar{\theta}$$  \quad (20)$$

This exactly mirrors Eq. (5), to the point that we use the same symbol $\bar{\theta}$ because these terms will turn out to be identical. For both a linear SAR and an ideal solution, the cross interaction term is simply the average of the diagonal terms. This is fundamental. There is a one-to-one correspondence between a linear SAR and ideal solution behavior, provided that the ideal solution is defined consistently with the SAR – not in terms of the molarity of the solute but in terms of the molarity of the functionalities comprising the solute and the mean-field properties of the solvent.

For compounds with similar densities in the OA system, the appropriate fraction $f_{i}^{M}$ is close to the volume fraction of the molecule $i$ in the OA solvent $s$. In practical terms this means that a linear SAR corresponds to ideal solution behavior on a mass basis, because the carbon and oxygen moieties all have very similar masses. Formulating the thermodynamic interactions with different fractions (i.e., mole, volume, mass, etc.) simply requires transforming the $x$-axis in Fig. 2. We are making no effort to exactly describe activity curves in that figure for individual molecules but rather seeking to develop reasonably accurate values for bulk OA and various measurable fractions, all with unknown molecular composition. Given roughly constant density, switching between volume and mass fractions amounts to a slight skewing of the $x$-axis in Fig. 2 that will fall out with empirical calibration along with the other simplifying assumptions made so far. This is consistent with the mass-fraction based solution formulation in the 1-D VBS (Donahue et al., 2006).
3.3.1 Non-linear SAR

If a linear SAR is equivalent to Eq. (20), a non-linear SAR must be given by \( \theta_{CO} \neq \bar{\theta} \). There are obviously many ways of introducing non-linearity, but we shall show that this method introduces both in a self-consistent and useful way suitable for real OA mixtures. Given the one-to-one relationship between \( lC^0 \) and \( \theta \) already developed, we can write a non-linear SAR in concise form as a matrix equation, adding a non-linear cross term \( \delta \theta_{CO} \) to the linear term introduced in Eq. (20):

\[
\theta_{i,s} = n_M^i \left( f^i_C f^i_O \right) \left[ \left( \frac{\theta_{CC}}{\bar{\theta} + \delta \theta_{CO}} \theta_{OO} \right) \left( f^S_C f^S_O \right) \right]
\]

or, for a single-component system, the pure, sub-cooled liquid solvation energy (and thus \( lC^0 \) SAR) is given by:

\[
\theta_{i,i} = n_C^i \theta_{CC} + n_O^i \theta_{OO} + 2 \frac{n_C^i n_O^i}{n_C^i + n_O^i} \delta \theta_{CO}
\] (22)

We still assume a linear relationship between the solvation energy of a pure substance given by Eq. (22) and \( lC^0 \), with

\[
lC^0_i = \left( \theta^0 - \theta_i \right) / \delta \theta_{10}
\]

or, equivalently

\[
lC^0_i = \left( n_C^0 - n_C^i \right) b_C - n_O^i b_O - 2 \frac{n_C^i n_O^i}{n_C^i + n_O^i} b_{CO}
\] (24)

As we shall see the cross term is sufficient to provide interesting and realistic behavior – and we can constrain it.

The non-linear SAR proposed here contains four free parameters:
1. The carbon-carbon interaction term \( \theta_{CC} = \delta \theta_{10} \cdot b_C \),
2. The oxygen-oxygen interaction term \( \theta_{OO} = \delta \theta_{10} \cdot b_O \),
3. The carbon-oxygen nonideality \( \delta \theta_{CO} = \delta \theta_{10} \cdot b_{CO} \), and
4. The solvation energy of a 1 \( \mu g m^{-3} \) substance \( \theta^0 = \delta \theta_{10} \cdot b_C \cdot n_C^0 \).

Our argument is that these four parameters underpin a reasonably accurate (and testable) thermodynamic framework. More information can be added – for example splitting the “oxygen” moiety into =O and -OH would add another diagonal and two new cross terms, and adding a nitrate moiety will require a similar expansion. We will certainly need to add the nitrate term for high-NO\(_x\) environments, but it remains to be seen whether the oxygen term will need to be split for realistic atmospheric situations.

We can put some real numbers on the interaction matrix given the data presented so far. We will use \( \delta \theta_{10} = 690 \text{ K} \), as discussed earlier. With this and the relations from Fig. 3, we have \( b_C = 0.475 \rightarrow \theta_{CC} = 328 \text{ K} \), and we could also use \( b_O = 1.7 \) to give \( \theta_{OO} = 1172 \text{ K} \). However, we shall consider this together with the non-linear cross term. The data shown in Fig. 3 provide much better constraints on the volatility of a homologous sequence of compounds (i.e. increasing \( n_C \) with constant functionality). In addition, most of the molecules constraining the solvation energy have many more carbon than oxygen atoms. Consequently, we do not really constrain \( \theta_{OO} \) as much as \( \partial \theta / \partial n_O \), generally with \( n_O < n_C \):

\[
\frac{\partial \theta_{i,i}}{\partial n_O} \approx \theta_{OO} + 2 \delta \theta_{CO} = 1172 \text{K}
\]  

(25)

Thus, in the linear limit \( \bar{\theta} = 750 \text{ K} \); this is equivalent to a linear \( \theta_{OO} \) of 1172 K. For the non-linear case we add \( \delta \theta_{CO} \) to the diagonal term instead of the off-diagonal term:

\[
\theta_{OO} = 1172 \text{K} - 2 \delta \theta_{CO}
\]  

(26)
This will keep the effect of oxygenated functionality consistent for relatively small \( n_\text{O} \) were the data constraints are robust, resulting in the interaction matrix:

\[
\theta = \begin{pmatrix}
328 & 750 \\
750 & 1172 - 2\delta\theta_\text{CO}
\end{pmatrix} K
\] (27)

The next question is, what is a reasonable non-linearity for typical ambient OA constituents? It is almost certain that \( \delta\theta_\text{CO} < 0 \) (the cross term is destabilizing) because this is what drives phase separation and we well know that oil and highly polar organics do not mix. For a constraint, we shall return to the volatility data presented in Fig. 3, noting two features. Most importantly, the diacids are less volatile than our simple linear SAR suggests, even though the same rules do a good job predicting mono-substituted organic acids. In addition, many of the sequences of oxygenated organics show a “hook” at low \( n_\text{C} \), with lower volatility in the data than predicted by the simple linear trend lines in Fig. 3 (for example, the alcohols and mono-acids for \( n_\text{C} < 4 \)). In Fig. 7 we show these data again, but now employing the non-linear SAR described by Eq. (24) with \( \delta\theta_\text{CO} = -0.3 \times 690 = -207K \). The consequence is an oxygen interaction term \( b_\text{O} = 2.3 \) but with \( b_\text{CO} = -0.3 \) generally compensating for \( n_\text{O} \ll n_\text{C} \) (i.e., low O:C). The new SAR, again summarized with dashed lines, gives better agreement with the data. Specifically, the volatility of the diacids is now reproduced with good fidelity, and the SAR (the thick dashed lines) shows a low-\( n_\text{C} \) hook.

The revised SAR in our 2-D space is displayed in Fig. 8. This is similar to Fig. 5, but close inspection reveals that the isopleths along the top line (O:C = 1:1) have shifted toward lower volatility by about 1 decade. For example, a volatility of 100 \( \mu \text{gm}^{-3} \) is appropriate for butanedioic (succinic) acid. The carbon isopleths are very slightly curved. This slight curvature toward lower volatility is crucial to the activity coefficients that drive phase separation between non-polar and polar organics. This is important, as it means that highly oxygenated organics of a given volatility will tend to have a somewhat lower carbon number than they would in a linear system. Also, by comparing Figs. 7 and 8, one can see that the region of this 2-D space that is reasonably constrained by
data extends to roughly the \( n_O = 4 \) green isopleth, which curves from \( n_C = n_O = 4 \) at \( \log_{10} C^O \approx 2 \) downward and leftward. This approximates the diacids. The space to the upper left of this curve the SAR is an extrapolation, and that is the very space occupied by ambient oxygenated organic aerosol, indicated by the dashed oval labeled “OOA” (Jimenez et al., 2009).

While the \( n_C, n_H \) and \( n_O \) isopleths in the OOA region in Fig. 8 are extrapolations, they are not extreme extrapolations. We consider this to be a reasonable projection of typical values for ambient organic-aerosol compounds, for all of the reasons discussed so far. Again, we emphasize that the primary quantities – \( \log_{10} C^* \) and O:C, can be constrained by ambient observations, and our primary application of the non-linear SAR is as a diagnostic to ascertain the typical values for \( n_C, n_H \) and \( n_O \).

4 OA activity coefficients

The SAR so far has focused on the pure solvation energies of molecules (really pseudo molecules) composed of carbon and oxygen groups. When we relax our conditions and allow a molecule \( i \) to interact with an OA solvent \( s \) with a different average composition, we wish to calculate the effective saturation concentration including an activity coefficient, \( C^* = \gamma C^O \). This will allow us to treat (model, not just diagnose) realistic OA mixtures with atmospherically relevant composition. Because in our volatility-based space we specify the saturation concentration \( /C^O \), we need to know the activity coefficient. As we showed in Eq. (9), this depends on the non-ideal off-diagonal interaction term. This is readily determined in our homogeneous solvent:

\[
\delta \theta_{i,s} = n_M \left[ \left( f_{ic}^i \right)^2 + \left( f_{sc}^s \right)^2 - 2 f_{ic}^i f_{sc}^s \right] \delta \theta_{CO}
\]  

(28)

This couples the non-ideal solution and non-linear SAR treatments we have developed so far. The non-linearity of the SAR (\( \delta \theta_{CO} \)) leads directly to any non-ideality of a solute in the OA solution. At this level of approximation, all of the non-ideality
rests on differences between the carbon fraction \( f_C \), or O:C in the solvent and solute. This is a substantial simplification; however, it amounts to asserting that polarity is the most important determinant of activity coefficients in organic solutions, which is justified (Prausnitz et al., 1998; Pankow and Barsanti, 2009). Furthermore, the approximate molar volume \( n_M \) of the solute molecule is also a critical parameter.

### 4.0.2 Activity coefficients with a non-linear SAR

To illustrate the consequences of non-ideality, we shall consider two fairly typical situations typified by OA with two different bulk (single-phase) compositions. First we shall consider a highly aged background aerosol with O:C = 0.75:1. This is consistent with highly oxidized OA characteristic of the remote atmosphere and generally called OOA (oxidized organic aerosol) (Jimenez et al., 2009; Ng et al., 2009). Second, we shall consider a much less oxidized aerosol with O:C = 0.25:1. This is more consistent with very fresh emissions, possibly combining traffic emissions (known as HOA, or hydrocarbon like organic aerosol to the AMS community) and either some biomass burning OA (BBOA) or slightly oxidized first-generation SOA (Lanz et al., 2007; Jimenez et al., 2009).

In Fig. 9 we show organic activity coefficients \( \gamma_{i,s} \) for the full range of organic solutes dissolved in these two different bulk aerosol solvents. This is calculated by using the nonlinear energy from Eq. (28) in Eq. (9) at 300 K using \( \delta \theta_{CO} = -207 \) K. The O:C of the bulk solvent is indicated on each panel. The activity coefficients of organics are contoured in these panels for all \( C^O \), O:C combinations, with contour intervals of 0.5 ranging from 1 to 10, labeled at integer values from 1 to 5. For example material at \( \log_{10} C^O = -5 \), O:C \( \approx 0.25:1 \) in Fig. 9a has \( \gamma \approx 10 \), resulting in \( \log_{10} C^* \approx -4 \). This model is highly simplified, but it contains essential features that can be constrained by ambient observations and that exhibit the fundamental characteristics of ambient OA.

In practical terms, when \( \gamma_{i,s} > 5 \) or 10, it is likely that a constituent will phase separate, as the activity \( a_i \) will be greater than 1.0 when \( f_i > 0.2 \) or 0.1. If there is any significant amount of a group of similar constituents, they will form their own phase.
Once a class is less than 10 or 20% of the OA, it is of secondary concern in any event. In AMS jargon, this implies that fresh HOA will not mix with aged OOA when OOA is in the majority (this is the “phase separation region” indicated along the lower edge of Fig. 9a). Note again that this fraction is the volume fraction, which is effectively the mass fraction, and we are judging ideality and modeling mixing on a volume-fraction basis.

What Fig. 9a shows is that when the background aerosol is highly oxygenated (with bulk average O:C indicated by the dark green arrow), non-polar, reduced primary emissions (POA with O:C ≃ 0 and −2 ≤ /C^0 ≤ 2, for example the gray dot) will have a hard time mixing into that background aerosol and will instead tend to phase separate (more accurately they will tend to remain in their original phase, often in their own size mode – for example a traffic mode). The phase boundary will be somewhere near the γ = 5 or 10 contours, depending on the exact composition distribution; the uncontoured region to the lower left of that boundary (low /C^0, low O:C) will form a separate, “POA” phase. The transition is quite sharp and the location is not very sensitive to the exact value of the off diagonal term δθ_{CO}. This is very consistent with much more sophisticated calculations for far simpler mixtures designed to simulate OA. For example, (Jang et al., 1997) showed that larger alkanes (exclusively) showed substantial activity coefficients in model mixtures designed to simulate both woodsmoke and SOA.

On the other hand, when the bulk OA is not very oxidized, as in Fig. 9b, almost all OA compounds can mix with the aerosol with only modest activity penalties (γ < 3). Only for semi-volatile compounds just at the volatility threshold (C_i^* ≃ C_{OA}) will the small difference in volatility due to the non-ideality have any consequence. These select compounds will shift toward the vapor phase because C_i^* > C^0 for γ > 1. For the most part, though, when the bulk OA is relatively fresh, with low O:C, we expect it to be characterized by a single condensed phase, which will likely become progressively more oxidized as the mixture ages (Jimenez et al., 2009). In AMS jargon, OOA will mix with HOA (or especially BBOA) when HOA is in the majority.
In general, the consequence of incorporating activity coefficients into any representation of OA will be to force some compounds into the vapor phase. This may become even more pronounced considering the dynamics of fresh mobile-source emissions, which can be small enough for the Kelvin term to increase the volatility even more (Zhang et al., 2004). This will be most dramatic when the systems phase separate, and the region where a separate phase may have a relatively small mass fraction is the lower left-hand portion of the 2-D space, which comprises fresh primary emissions (POA, or HOA in AMS jargon).

However, it would be a gross oversimplification to conclude that incorporation of activity coefficients into a model coupling aerosol dynamics and chemistry would result in lowered OA levels due to the increased volatility. The consequence of forcing more HOA into the gas phase will almost certainly be rapid gas-phase oxidation (Shrivastava et al., 2008), generating products that will mix with the OOA phase with mass yields that may even exceed unity (Presto et al., 2009, 2010). The issue is thus not whether POA can act as a seed for SOA, but rather whether SOA (really OOA) is a good seed for POA. The answer is no, but the full answer is not that there will be less OA, but rather that POA may for a time exist in a separate, fresh size mode but that it will tend to evaporate, be more rapidly oxidized in the gas phase, and generate oxidation products that will mix into SOA. Thus the major observable consequence of non-ideality may be kinetic – the transformation from POA to SOA, or HOA to OOA, will be more rapid that it would otherwise be were an ideal solution assumed.

5 Conclusions

We have presented a self-consistent thermodynamic foundation for a two-dimensional volatility basis set relying on a minimum number of free parameters to describe realistic and measurable properties of real-world organic-aerosol mixtures. This treatment suggests that we can reasonably describe OA thermodynamics with a mean-field approximation including a non-ideality term that simultaneously and self-consistently de-
scribes non-linearity in the structure activity relation (SAR) as well as non-ideality in the solution thermodynamics. Because $\delta \theta_{CO} < 0$, the consequences of all this will be that more molecules will find themselves in the gas phase than in an ideal world.

The formulation so far does not address the role of water (Pankow and Chang, 2008; Tong et al., 2008; Barley et al., 2009; Prisle et al., 2010), but the strong correlation of O:C and OA hygroscopicity reported in Jimenez et al. (2009) suggests that we may be able to treat water in a similar fashion. However, because water can be a dominant component, and therefore its unique properties are more likely to challenge the mean-field assumption, it is likely that explicit treatment of humidity effects will be required (Prisle et al., 2010).

The 2-D-VBS complements our well-established 1-D-VBS (Donahue et al., 2006) by adding oxygen content (O:C) as a second dimension in addition to saturation concentration ($C^o$). Using pairwise interaction energies we can formulate both a simple structure activity relation to relate carbon and oxygen numbers to O:C and $C^o$. The formulation includes a non-linear coupling term that can be constrained by pure component vapor-pressure data that also leads to realistic activity coefficient predictions. This formulation relies on measurable bulk properties of organic aerosol (most notably O:C) and uses a mean-field approximation for the molecule-solvent interactions – in essence we assume that a molecule interacts more with other functional groups in the organic aerosol solution than with other molecules in that mixture. Importantly, the complex behaviors develop from a very simple formulation containing only four free parameters, all of which are constrained by vapor-pressure data of pure substances.

This is a foundation for future work. In several succeeding publications we shall explore the uses of this 2-D-VBS for both diagnostic analyses and prognostic calculations organic aerosol properties, ultimately extending the model to a fully coupled treatment of size-dependent organic-aerosol dynamics.
Acknowledgements. This work was funded by the US EPA Science to Achieve Results (STAR) program.

References


and vapor pressures of very low volatility compounds: A study of the C-4-C-10 and C-12
behavior of a mixture of organic aerosol components, P. Natl. Acad. Sci. USA, 105, 18 687–
18 691, doi:10.1073/pnas.0802144105, 2008. 24100, 24102
Carlton, A. G., Turpin, B. J., Lim, H.-J., Altieri, K. E., and Seitzinger, S.: Link between isoprene
and secondary organic aerosol (SOA): Pyruvic acid oxidation yields low volatility organic
Chandramouli, B., Jang, M., and Kamens, R.: Gas-particle partitioning of semi-volatile or-
ganics on organic aerosols using a predictive activity coefficient model: analysis of the
effects of parameter choices on model performance, Atmos. Environ., 37, 853–864, doi:
10.1016/S1352-2310(02)00931-7, 2003. 24094
Clegg, S. L., Kleeman, M. J., Griffin, R. J., and Seinfeld, J. H.: Effects of uncertainties in the
thermodynamic properties of aerosol components in an air quality model Part 1: Treatment
of inorganic electrolytes and organic compounds in the condensed phase, Atmos. Chem.
Clegg, S. L., Kleeman, M. J., Griffin, R. J., and Seinfeld, J. H.: Effects of uncertainties in the
thermodynamic properties of aerosol components in an air quality model - Part 2: Predic-
tions of the vapour pressures of organic compounds, Atmos. Chem. Phys., 8, 1087–1103,
doi:10.5194/acp-8-1087-2008, 2008. 24094
Donahue, N. M. and Prinn, R. G.: In-situ nonmethane hydrocarbon measurements on SAGA-3,
Donahue, N. M., Robinson, A. L., Stanier, C. O., and Pandis, S. N.: Coupled partitioning,
dilution, and chemical aging of semivolatile organics, Environ. Sci. Technol., 40, 2635–2643,
2006. 24093, 24097, 24106, 24109, 24117
Donahue, N. M., Robinson, A. L., and Pandis, S. N.: Atmospheric Organic Particulate Matter:
Epstein, S. A., Riipinen, I., and Donahue, N. M.: A Semi-Empirical Correlation between En-
thalpy of Vaporization and Saturation Concentration for Organic Aerosol, Environ. Sci. Tech-
2010. 24096, 24097
Fraser, M. P., Cass, G. R., Simoneit, B. R. T., and Rasmussen, R. A.: Air quality model eval-
24119


Fig. 1. Solvation energy conceptualized. A molecule, $i$ (teal circle), is dissolved in a solvent $s$. To evaporate it must be removed from the solvent, leaving a void and costing energy $\theta_{i,s}$ to both the molecule and the solvent, but the solvent will relax to fill the void, releasing energy $\theta_{s,s}$. 

$$f_C, f_O = (1 - f_C)$$
Fig. 2. Activity of component A at 300 K for two-component (A-B) mixtures with non-ideal interaction energies $\delta \theta_{AB}$ of 200, 300, 400, and 500 K. For $\delta \theta > 300$ the system rapidly phase separates, quickly achieving almost complete separation.
Fig. 3. Structure-activity trends for singly and doubly functionalized oxygenated organics. Volatility \( \left( \log_{10} C^o \right) \) vs. \( n_C \) (carbon number) for different classes of organic compounds (indicated with colored symbols) forming the basis for the carbon and oxygen isopleths presented in the top panel. Each class contains the same oxygen functionality but a range of \( n_C \). The slopes of the class lines show the effect of increasing \( n_C \) \((-0.475 \) decades per C\), while offsets from the hydrocarbon line show the effects of functionalization. Carbonyls decrease \( C^o \) by about 1 decade per (=O) while alcohols decrease \( C^o \) by about 2.3 decades per (-OH). For a linear approximation we adopt an average of \(-1.7 \) decades per O, indicated with dashed lines for 0, 1, 2, and 4 oxygens; later (Fig. 7) we shall add a modest nonlinearity.
Fig. 4. The effect of different functional groups on volatility ($C^o$) and O:C for a C$_6$ and C$_{20}$ backbone. HO- is highly polar, decreasing $/C^o$ by 2.3 per group, leading to the shallow lines. O= is less polar, decreasing $/C^o$ by 1 per group, leading to steeper lines. Acids (HO(O)-) decrease $/C^o$ by 1.75 per group. The average effect of oxygentated functionality deduced in this work (in black) leads to slopes similar to the structure-activity of organic acids.
Fig. 5. Composition based on a simple, linear structure activity relation described in the text, displayed in a 2-D space with volatility ($\log_{10} C^o$) on the x-axis and oxygenation (O:C) on the y-axis. Isopleths are for constant carbon number ($n_C$, labeled black lines), oxygen number ($n_O$, green curves) and total functional group number ($n_M = n_C + n_O$, cyan curves). The rough range of oxygenated organic aerosol (OOA) seen in the atmosphere is shown with a dashed oval.
Fig. 6. Solvation interactions in a simple Structure Activity Relation (SAR) for organics dissolved in an organic-aerosol solvent.
Fig. 7. Volatility ($C^o$) predictions with modest nonlinearity ($b_{CO} = -0.3$) in a structure-activity relation. As in Fig. 3, the average effect of oxygenated functionality is shown by thick dashed curves. The consequence of the nonlinearity is to increase the interaction energies of oxygenated functional groups, leading to lower volatility of diacids and a pronounced hook at low $n_C$ (and thus a high oxygen fraction) – these are both consistent with the data (e.g. alcohols and acids for $n_C < 4$).
Fig. 8. Organic composition with a small degree of non-linearity (−0.3 decades per C-O interaction) in the SAR describing OA volatility. Isopleths are for of constant carbon number ($n_C$, labeled black lines), oxygen number ($n_O$, green curves) and total functional group number ($n_M = n_C + n_O$, cyan curves). All isopleths are for even numbers only, and the values for the green and cyan isopleths can be determined by the intersection points at O:C = 1:1 (e.g. the meeting of $n_C = n_O = 4$ at $\log_{10}(C^o) = 2$ and O:C = 1, with $n_M = 8$). Compared with Figure 5 more oxygenated organics are somewhat less volatile – the isopleths are shifted approximately 1 decade to the left for O:C = 1:1.
Fig. 9. Activity coefficients of organics in the 2-D space for two different mean OA compositions. In the top panel the mean O:C of the bulk OA is 0.75:1 (OOA-like), indicated by the thick green arrow, while in the bottom panel the mean O:C in the bulk OA is 0.25:1 (POA-like), indicated by the thick gray arrow. The isopleths are for $\gamma_{i,s}$, the activity coefficient of organics with the $IC^*$ and O:C indicated in each solution. Contours are steps of 0.5 in $\gamma$, starting at 1 and ending at 10. Integer values are labeled up to $\gamma = 5$. Somewhere between $5 < \gamma < 10$ phase separation becomes very likely. For example, the POA constituent indicated by the dark gray circle in in top panel at $\log_{10} C^o = -2$, O:C = 0:0 has an activity coefficient in the bulk OOA $\gg 10$; conversely, the OOA constituent indicated by the green circle in in bottom panel at $\log_{10} C^o = -2$, O:C = 0.75:1 has an activity coefficient in the bulk OOA $= 2.5$. 