Supplemental Information for:
RH-dependent organic aerosol thermodynamics via an efficient reduced-complexity framework

Kyle Gorkowski[1], Thomas C. Preston[1, 2], and Andreas Zuend[1]

[1]Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada
[2]Department of Chemistry, McGill University, Montreal, Quebec, Canada

Correspondence: Kyle Gorkowski (kyle.gorkowski@mcgill.ca), Andreas Zuend (andreas.zuend@mcgill.ca)

1 Overview

The supplemental information covers the BAT model equations and the approaches for the parameterizations of different functional group classes and phase separation treatments. These approaches include the O : C blending method developed for the transition regions between the three BAT model parameterization regions, the functional group translations approach to convert input parameters to OH-group equivalents, finding the $a_{w,\mathrm{sep}}$ point for the liquid–liquid transition from a organic-rich to a water-rich phase, and the density estimation method for organic compounds. The attached supplemental Microsoft® Excel workbook file contains all the coefficient values, the SOA model system's input properties, validation systems, and all the data shown in the figures of the main text.

2 BAT model

2.1 BAT Equations

The explicit equations for our BAT model are listed below in Eqs. (S1) to (S11). To improve the clarity, we define $\mathrm{O:C} \equiv \vartheta$, where O : C refers to the O : C of an organic component ("org") or the average O : C of a mixture of organics. The determined coefficients are listed in Tables S1 & S2.

$$c_1 = a_{1,1}\ \exp(a_{1,2}\ \vartheta) + a_{1,3}\ \exp\left(a_{1,4}\ \frac{M_w}{M_{org}}\right) \tag{S1}$$

$$c_2 = a_{2,1}\ \exp(a_{2,2}\ \vartheta) + a_{2,3}\ \exp\left(a_{2,4}\ \frac{M_w}{M_{org}}\right) \tag{S2}$$

$$\phi_{org} = x_{org}\left(x_{org} + (1 - x_{org})\frac{\rho_{org}}{\rho_w}\frac{M_w}{M_{org}}\left[s_1(1+\vartheta)^{s_2}\right]\right)^{-1} \tag{S3}$$

$$G^E/RT = \phi_{org}(1-\phi_{org})\left[c_1 + c_2(1-2\phi_{org})\right] \tag{S4}$$

$$\frac{d(G^E/RT)}{dx_{org}} = \frac{d(G^E/RT)}{d\phi_{org}}\frac{d\phi_{org}}{dx_{org}} \tag{S5}$$

$$\frac{d\phi_{org}}{dx_{org}} = \left(\frac{\rho_{org}}{\rho_w}\frac{M_w}{M_{org}}\left[s_1(1+\vartheta)^{s_2}\right]\right)\left(\frac{\phi_{org}}{x_{org}}\right)^2 \tag{S6}$$

$$\frac{d(G^E/RT)}{dx_{org}} = \left\{(1-2\phi_{org})\left[c_1 + c_2(1-2\phi_{org})\right] - 2c_2\phi_{org}(1-\phi_{org})\right\}\frac{d\phi_{org}}{dx_{org}} \tag{S7}$$

$$\ln(\gamma_{org}) = (G^E/RT) + (1-x_{org})\frac{d(G^E/RT)}{dx_{org}} \tag{S8}$$

$$a_{org} = \gamma_{org}x_{org} \tag{S9}$$

$$\ln(\gamma_w) = (G^E/RT) - x_{org}\frac{d(G^E/RT)}{dx_{org}} \tag{S10}$$

$$a_w = \gamma_w(1-x_{org}) \tag{S11}$$

Here, the activity coefficients of organic and water, γ_{org} and γ_w, respectively, as well as the corresponding activities (a_{org}, a_w) are defined on mole fraction basis (i.e. γ_{org} = $\gamma_{org}^{(x)}$), each with the pure component as reference and standard states (where activity coefficients become unity). The output from the BAT calculation can also be used to calculate the Gibbs energy of mixing ($\Delta_{\mathrm{mix}}G$), since the non-ideal interactions are parameterized (i.e., the excess Gibbs energy of mixing: G^{E}). Note, for simplicity, we do not include standard state chemical potentials of water and the organic, which would add an additional linear component to the curve. This is deemed justified given the approximate nature of the miscibility gap treatment. We present this calculation below with $\Delta_{\mathrm{mix}}G$ being normalized by R, T, and the total sum of moles $n_t = n_w + n_{org}$ in the binary system.

$$\frac{\Delta_{\mathrm{mix}}G^{\mathrm{ideal}}}{RTn_t} = (1-x_{org})\ln(1-x_{org}) + x_{org}\ln(x_{org}) \tag{S12}$$

$$\frac{\Delta_{\mathrm{mix}}G}{RTn_t} = \frac{\Delta_{\mathrm{mix}}G^{\mathrm{ideal}}}{RTn_t} + \frac{G^E}{RTn_t} \tag{S13}$$

Table S1. Scaled volume coefficients of the fitted BAT model.

Region	O : C bounds	Training data points	s_2	s_1
low O : C	$\mathrm{O:C} < 0.15$	1000	-5.988895	6.940689
mid. O : C	$0.05 < \mathrm{O:C} < \vartheta_{\mathrm{ML}} + 0.1$	2680	-1.219164	4.742729
high O : C	$\vartheta_{\mathrm{ML}} < \mathrm{O:C}$	3600	-0.078682	3.650860
misciblity line	$0.05 < \mathrm{O:C} < 0.45$	2360	-1.237227	4.069905

Table S2. The eight power series coefficients ($a_{n,1-4}$; $n = 1, 2$) used in the hydroxyl-group-parameterized BAT model.

Region	$a_{1,1}$	$a_{2,1}$	$a_{1,2}$	$a_{2,2}$	$a_{1,3}$	$a_{2,3}$	$a_{1,4}$	$a_{2,4}$
low O : C	7.089476	-0.622678	-7.711860	-100.0	-38.859410	3.08E-09	-100.0	61.888120
mid. O : C	5.872214	-0.974049	-4.535007	-100.0	-5.129327	2.109751	-28.092320	-23.676830
high O : C	5.921550	-100.0	-2.528295	-100.0	-3.883017	1.353916	-7.898128	-11.601450
misciblity line	5.885109	-0.984901	-4.731250	-6.227207	-5.201652	2.320286	-30.822970	-25.840370

2.2 Limit of Miscibility Line

The limit of miscibility line is determined from an initial BAT model fit involving the O : C region close to where the miscibility gap *vs.* complete miscibility transition occurs. We started by fitting the BAT coefficients using a wide O : C range (0.0 to 0.8) and then progressively narrowed it to the transition region (O : C 0.05 to 0.45). We then scanned O : C and M_{org} to map out where the miscibility transition occurred (within BAT). The resulting O : C values were used to fit the limit of miscibility line, ϑ_{ML}, as a function of organic molar mass,

$$\vartheta_{\mathrm{ML}} = \frac{0.205}{1 + \exp\left(26.6\left(\frac{M_w}{M_{org}} - 0.12\right)\right)^{0.843}} + 0.23. \tag{S14}$$

2.3 O : C Transition Region Blending

We used three different sets of fitted coefficients for the base BAT model representing hydroxyl functionality molecules. The split was based on the limit of complete miscibility of organics with water and further separated by O : C. A sigmoidal function was introduced to provide a smooth transition when traversing from one of the domains to the next in the 2-D parameter space (e.g., when O : C is increased gradually at a constant molar mass coordinate) – otherwise, spurious discontinuities would occur. The sigmoidal function provides a weighted map between the parameters from one domain to the next (over a short range in the boundary region). In effect, we are blending the different regions in the hydroxyl BAT model. Low to medium O : C region blending is listed first (Eqs. S15 to S22), where ϑ_{ML} is the ϑ value at the limit of miscibility line and b_1, b_2, and b_{ML} are the

blending coefficients (Table S3). These are followed by and example of the blending weights as a function of O : C, Fig. S1.

$$\vartheta_b = \vartheta - \vartheta_{\mathrm{ML}} b_{\mathrm{ML}} \tag{S15}$$

$$\varpi_b = \frac{1}{1+\exp[-b_1(\vartheta_b - b_2)]} \tag{S16}$$

$$\vartheta_{b,norm} = \vartheta - 0.75\,\vartheta_{\mathrm{ML}}\, b_{\mathrm{ML}} \tag{S17}$$

$$\varpi_{norm} = \frac{1}{1+\exp(-b_1(\vartheta_{b,norm} - b_2))} \tag{S18}$$

$$\varpi_{mid} = \varpi_b / \varpi_{norm} \tag{S19}$$

$$\varpi_{low} = 1 - \varpi_{mid} \tag{S20}$$

$$G^E/RT\Big|_{blended} = \varpi_{low}\, G^E/RT\Big|_{low} + \varpi_{mid} G^E/RT\Big|_{mid} \tag{S21}$$

$$\frac{d(G^E/RT)}{dx_{org}}\Big|_{blended} = \varpi_{low}\, \frac{d(G^E/RT)}{dx_{org}}\Big|_{low} + \varpi_{mid}\, \frac{d(G^E/RT)}{dx_{org}}\Big|_{mid} \tag{S22}$$

Medium to high O : C region blending (Eqs. S23 to S27):

$$\vartheta_b = \vartheta - \vartheta_{\mathrm{ML}} \tag{S23}$$

$$\varpi_{high} = \frac{1}{1+\exp(-b_1(\vartheta_b - b_2))} \tag{S24}$$

$$\varpi_{mid} = 1 - \varpi_{high} \tag{S25}$$

$$G^E/RT\Big|_{blended} = \varpi_{high} G^E/RT\Big|_{high} + \varpi_{mid} G^E/RT\Big|_{mid} \tag{S26}$$

$$\frac{d(G^E/RT)}{dx_{org}}\Big|_{blended} = \varpi_{high} \frac{d(G^E/RT)}{dx_{org}}\Big|_{high} + \varpi_{mid} \frac{d(G^E/RT)}{dx_{org}}\Big|_{mid}. \tag{S27}$$

Table S3. Coefficients used in the blending of the different BAT coefficient regions for a molecule with hydroxyl functionality.

Region Transition	b_1	b_2	b_{ML}
low to mid. O : C	79.2606902	6.04293E-02	0.1899745
mid. to high O : C	75.0159268	9.47111E-04	-

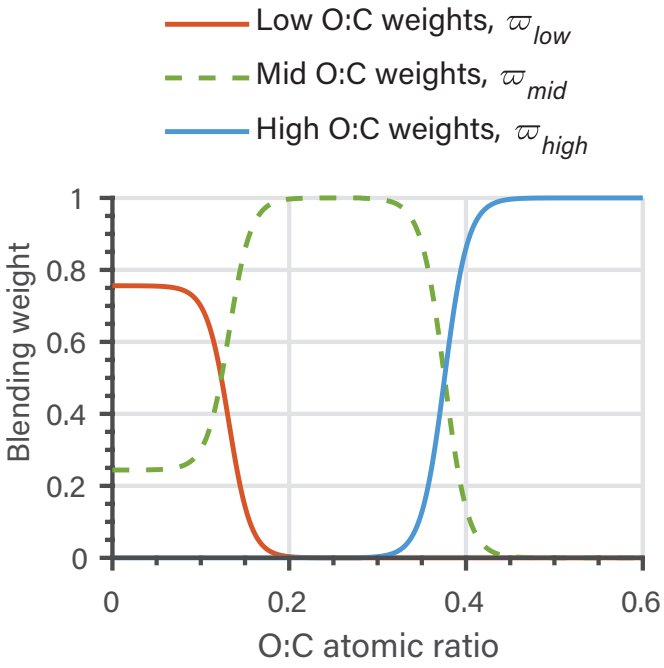

Figure S1. Example of the blending weights used to merge the three regions in the BAT model. The O : C is scanned with a fixed M_{org} of 200 g mol^{-1} to show how each region becomes dominant.

2.4 BAT Molecular Functionality Translation

The translation approach concerns the conversion from different molecular functionalities to hydroxyl-equivalent input parameters for use with the default, hydroxyl-group-based BAT model. These translations are for the whole molecule, and not the individual functional groups. Thus, for multifunctional molecules, a distinct multifunctional translation must be derived, as we did for the SOA oxidation products. If that is not possible, then the most dominant and representative functionality should be chosen. The O : C conversion is described by Eq. S28 and the molar mass translation is described by Eq. S29. The corresponding coefficients for different oxygen-bearing functionalities of the whole molecule are listed in Table S4.

$$\vartheta_{eqv.\mathrm{OH}} = \frac{\vartheta}{1 + t_3 \exp(-t_1\,\vartheta)} \tag{S28}$$

$$M_{eqv.\mathrm{OH}} = \frac{M}{1 + t_4 \exp(-t_2\,M)} \tag{S29}$$

Table S4. Functional group translation coefficients to convert a whole molecule to a hydroxyl-equivalent molecule for BAT model inputs.

t_n	Hydroxyl	Carboxyl	Hydroperoxide	Hydroperoxide SOA	PEG	Ketone	Ether	Ester
t_1	none	none	8.1716E-06	1.4902E-04	5.4477E-03	4.5343E-03	2.4434E-05	-1.293246
t_2	none	none	4.5318E-07	4.7363E-03	3.864336	6.4845E-04	1.5832E-04	1.0813E-03
t_3	none	none	0.966090	0.869058	-0.267168	0.138144	0.284974	1.240514
t_4	none	none	0.459433	0.564783	0.255487	0.352454	0.229339	0.405354

3 Water Activity Separation Point

In the case of a liquid–liquid equilibrium, the relative phase preferences are described by q_j^α, the fractional liquid–liquid partitioning of a component to phase α ($q_j^\alpha \leq 1.0$ in the two-liquid-phases case). Liquid–liquid phase separation (LLPS) in a binary water–organic system at RH $< 100\,\%$ is reduced to a point and manifests itself by a jump discontinuity. The liquid phase is either a water-poor (β) or water-rich (α) phase, with a sharp transition between these two possible states at a specific water activity ($q_j^\alpha = 1$ or 0). To approximate the location and a_w-width over which the liquid–liquid phase separation is prescribed to occur, we first determine a designated reference point, the so-called water activity separation point ($a_{w,\mathrm{sep}}$). Liquid–liquid phase separation connects two points on the Gibbs energy of mixing curve that have identical slopes and a tie-line that does not cross the Gibbs energy curve (Fig. S2a). This tie-line represents the connection between the two stable phase compositions at equilibrium. Prior to phase separation occurring, a mixture can enter the composition space past these two points, which will result in a metastable state and eventually an unstable state, which will lead to spontaneous, spinodal decomposition (if phase separation did not occur within the metastable region). The binary mixture can enter and remain in the metastable region, but the energy barrier for liquid–liquid phase separation is typically low at room temperature, such that phase separation is expected to occur when the water content is increased. In most cases we will be interested in a case of increasing or decreasing water mole fraction at approximately constant temperature, so our $a_{w,\mathrm{sep}}$ point in Fig. S2a will be p_2, which has a corresponding point p_5 near/within the metastable composition range. If we solved for the tie-lines at high precision and included the standard state chemical potentials of water and the organic, then points p_1 and p_2 would have identical activities. That however is not the case, but we still want to ensure identical water activities at $a_{w,\mathrm{sep}}$. We achieve this by finding p_2's corresponding point (p_5) which has the same water activity as the $a_{w,\mathrm{sep}}$ point, this ensures a realistic water-poor (β) to water-rich (α) transition.

Here, we explain how to identify (to good approximation) the two stable composition points in liquid–liquid equilibrium by only using the BAT-predicted activity curves (Fig. S2b). In a binary system, both component activities must be less than one and have monotonic behavior. Any regions that show non-monotonic behavior result in a phase separation range and are denoted by the dashed lines in Fig. S2b. By connecting the mole fraction extent of the organic and water activity-based (minimum) phase separation regions identified, we can construct the tie-line that connects the two stable phases over the full extent of phase separation. This tie-line is then used in our above description to find the $a_{w,\mathrm{sep}}$ point. We note that due to omitting a computationally costly Gibbs energy minimization (with further including standard chemical potentials), the identified miscibility gap is a (typically good) approximation of the true extent of phase separation.

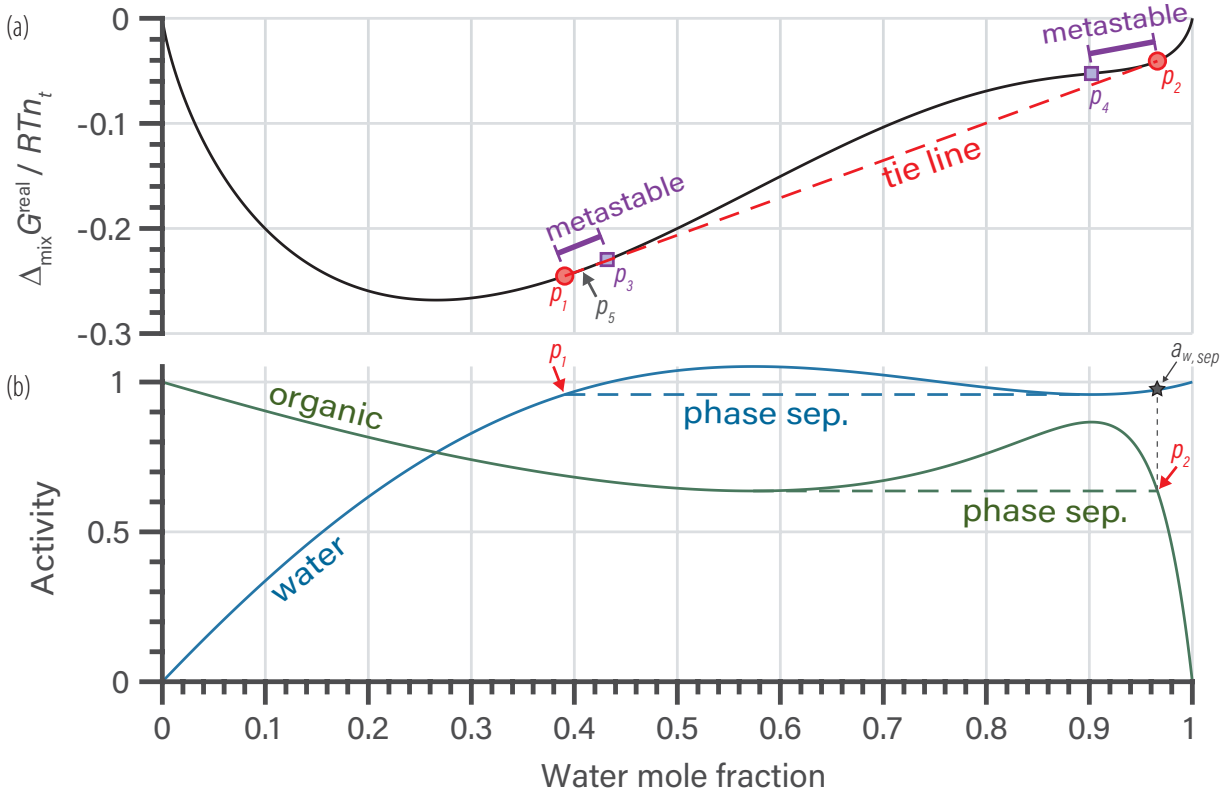

Figure S2. BAT simulation used to describe the identification of the $a_{w,\mathrm{sep}}$ point. The simulation uses an organic compound with hydroxyl functionalities, M_{org} of 100 g mol^{-1} and O : C of 0.225. The identified $a_{w,\mathrm{sep}}$ value is here 0.9741 (black star). (a) The normalized $\Delta_{\mathrm{mix}}G$ curve (black) with the tie-line in dashed red. The approximate stable phase-separation tie-line points and compositions are marked by p_1 and p_2, with the extent of the corresponding metastable regions denoted by p_3 and p_4. The end point in the metastable region at the same water activity as p_2 is marked by p_5. (b) The organic (green) and water (blue) mole-fraction-based activities for this binary system. The apparent minimum regions of phase separation required by each component are indicated by dashed lines. The approximate mole fraction extent of the actual phase separation region is identified by the extremes in composition, i.e. end points p_1 and p_2. The $a_{w,\mathrm{sep}}$ point is the water activity corresponding to the composition at p_2, indicated by a black star.

4 Organic Density Estimation

Organic density model from Girolami (1994), Eqs. (S34 to S36). If H : C is not known then we use $\mathrm{H:C} = 2 - \vartheta$.

$$M_C = 12.010\ \mathrm{g\,mol^{-1}} \tag{S30}$$

$$M_N = 14.006\ \mathrm{g\,mol^{-1}} \tag{S31}$$

$$M_O = 16.0\ \mathrm{g\,mol^{-1}} \tag{S32}$$

$$M_H = 1.008\ \mathrm{g\,mol^{-1}} \tag{S33}$$

$$n_c = \frac{M_{org}}{M_C + M_H\,\mathrm{H:C} + M_O\,\vartheta + M_H\,\mathrm{N:C}} \tag{S34}$$

$$\rho^* = \frac{M_{org}}{5n_c(2 + \mathrm{H:C} + 2\,\vartheta + 2\,\mathrm{N:C})} \tag{S35}$$

$$\rho_{est.} = \rho^*(1 + \min(0.1\,n_c\,\vartheta + 0.1\,n_c\,\mathrm{N:C},\ 0.3)) \tag{S36}$$

5 BAT Model Validation and Error Analysis

Given that the BAT model is a multivariate function, a validation data set is used to assess the possibility of overfitting of the model depending on the training data set. The species used in the training and validation (Table S6) data sets are listed in the attached MS Excel file, the summary of the error analyses are shown in Table S5. Figure S3 compares the calculated water and organic activities at the same organic mole fraction, which is clearer than directly comparing activity coefficients from each model. For O : C values lower than 0.2, the deviation from the 1:1 line is more substantial than the deviation for higher O : C compounds. This is expected as such compounds show a miscibility gap over a wide range of composition space and associated high activities when computed for the initial, well-mixed single-phase case. For a quantitative assessment we calculated the root mean squared error (RMSE) of the activities predicted by the two models (AIOMFAC being the benchmark). If there was substantial overfitting, there would be a large difference in the RMSE between the training data and the validation data. For the RMSE calculation, we excluded the points where the activity was greater than one, as those represent unstable physical states and large deviations there can overwhelm the RMSE. Model–model deviations for those unstable cases are largely irrelevant in practise, because what matters is the comparison of the predictive skill for the composition of the stable phases (in LLPS or single-phase case). Table S5 lists the compiled error assessments for the training data and the validation data. The similar RMSE values between the hydroxyl training and validation data suggest the model is not overfitting and has general applicability within the training domain of the parameter space (O : C and molar mass ranges). This agreement suggest that model behavior is realistic and our excess Gibbs function is smooth with no discontinuities. The smooth excess Gibbs function then leads to smooth activity curves and activity coefficients. Discontinuities like liquid–liquid phase separation are only derived from analysis of the excess Gibbs function (via post-processing) and are not directly built into the coefficients of the BAT model.

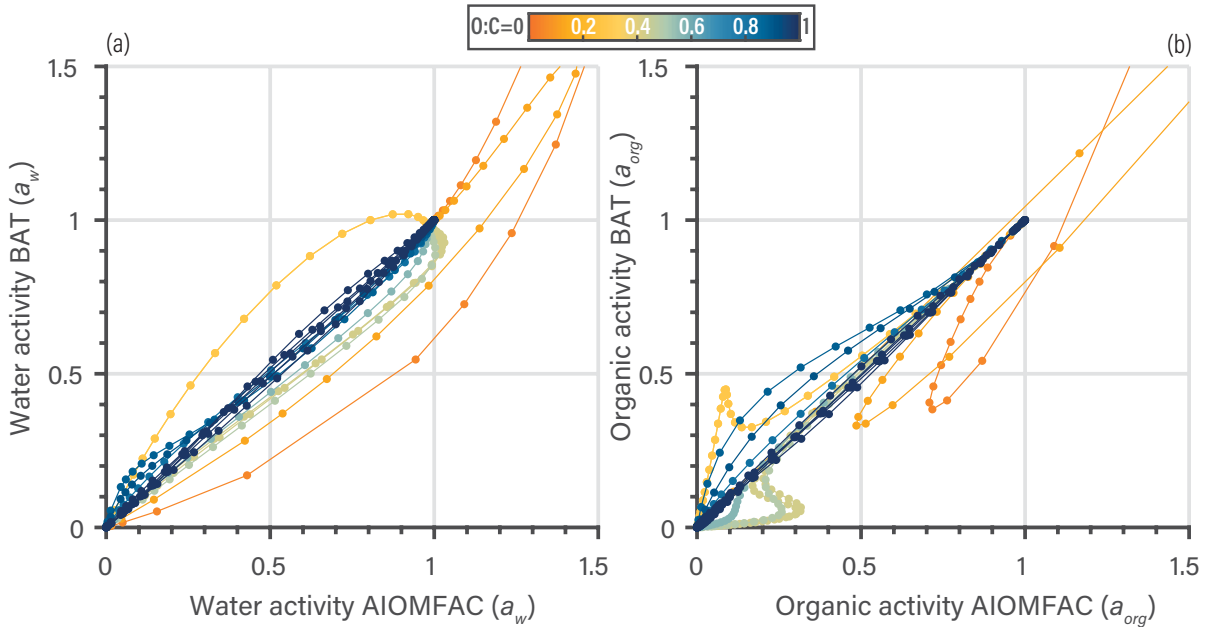

Figure S3. Comparisons of the BAT predictions with the AIOMFAC model predictions for the validation data set (Table S6). An activity value above one represents an unstable or metastable mixing state, and in practice the mixture would phase separate readily when given that initial mixture composition. The color bar represents the O : C of the compound, and each organic–water system is shown by 40 comparison points spanning the composition range from dilute to concentrated. Water activity ($a_w = \gamma_w(1 - x_{org})$) is shown in (a) and organic activity ($a_{org} = \gamma_{org} x_{org}$) in (b). Both models are compared at the same organic mole fraction, x_{org}.

We did not generate additional validation data sets for the translation coefficients for each molecular functionality type for two reasons. First, our translation has only four coefficients and will be well constrained by 100+ data points used in the fit. Second, our translation function constitutes a smooth map; thus, no artifacts due to potential overfitting are expected.

In addition to the thermodynamic activities, we can also compare how well we detect and predict the $a_{w,\mathrm{sep}}$ point. For the organic compounds in the binary aquous systems that underwent phase separation, the RMSE of BAT vs. AIOMFAC $a_{w,\mathrm{sep}}$ predictions are listed in Table S5. Overall the BAT $a_{w,\mathrm{sep}}$ prediction was $< \pm 0.01$, the a_w prediction was $< \pm 0.09$ (9 % RH), and the a_{org} prediction was $< \pm 0.15$ compared to AIOMFAC.

5.1 CCN Hygroscopicity Parameter Validation

We compare, in Fig. S4, measurement-derived κ_{CCN} data against the BAT and AIOMFAC model predictions of κ_{CCN}. The validation dataset contained 16 supersaturated growth measurements on known chemical species, listed in Table S6 (Petters et al., 2009; Broekhuizen et al., 2004; Brooks et al., 2004; Frosch et al., 2010; Huff Hartz et al., 2006; Petters et al., 2016, 2006; Pradeep Kumar et al., 2003; Raymond, 2003; Suda et al., 2014; Svenningsson et al., 2006). The average error in the

Table S5. BAT model data point numbers for model fit and validation as well as root mean squared errors (RMSE) for the training and validation databases, which were generated by the AIOFMAC model.

	Hydroxyl (training)	Hydroxyl (validation)	Carboxyl	Hydroperoxide	Hydroperoxide SOA	PEG	Ketone	Ether	Ester
Points for activity comparison ($a < 1$)	5511	607	451	573	910	120	421	557	488
RMSE of a_w	0.0580	0.0667	0.0408	0.0690	0.0711	0.0335	0.0845	0.0730	0.0820
RMSE of a_{org}	0.0901	0.0964	0.0771	0.0950	0.0982	0.0520	0.1320	0.0970	0.1450
Points for LLPS comparison	52	4	5	9	5	none	10	9	21
RMSE of $a_{w,\mathrm{sep}}$	0.0066	0.0127	0.0031	0.0039	0.0061	none	0.0075	0.0032	0.0024

measurements is shown as the shaded gray region and is the average of the κ_{CCN} range observed. The validation data shows similar agreement between the two models with a measurement vs. BAT RMSE of 0.061 and measurement vs. AIOMFAC RMSE of 0.059. The AIOMFAC κ_{CCN} predictions are better in the miscibility transition region than those from the BAT model, but overall both models show similar predictions.

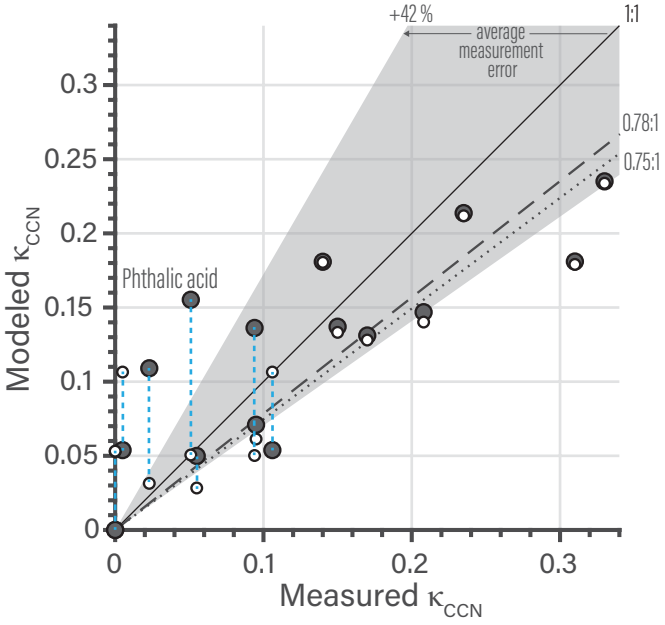

Figure S4. Comparison of single-component organics κ_{CCN} measurements against those predicted by BAT (black circles) and AIOMFAC (white circles) model simulations of CCN activation. The blue, dashed lines connect the BAT and AIOMFAC predictions for the same species when there is a large difference. Gray shading represents $\pm$ 42 % average uncertainty in the measured κ_{CCN}. The black dashed line is the BAT model linear fit with a zero intercept, $\kappa_{\mathrm{CCN,BAT}} = \kappa_{\mathrm{CCN,measured}} \times 0.78$ $[\pm 0.078]$ with a Pearson's $R^2 = 0.48$. The black dotted line is the AIOMFAC model linear fit with a zero intercept, $\kappa_{\mathrm{CCN,AIOMFAC}} = \kappa_{\mathrm{CCN,measured}} \times$ 0.75 $[\pm 0.066]$ with a $R^2 = 0.57$. The RMSE between the measurements and predictions were 0.061 for BAT and 0.059 for AIOMFAC. The simulations assumed a 100 nm diameter equivalent volume of organic matter at the CCN activation point and the droplet surface tension was calculated as a volume-weighted mean. A list of the 16 validation points is given in Table S6

Table S6: Chemical species used in the κ_{CCN} measurement comparison, which contains the subset of 16 species used for BAT model validation.

Start of Table S6									
Chemical	Validation Data	BAT functionality	O : C	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$)	BAT κ_{CCN}	AIOMFAC κ_{CCN}	Measured κ_{CCN}	Measurement Reference
Cetyl alcohol	yes	hydroxyl	0.06	2.00	242.50	1.93E-06	0.053	2.00E-05	Petters (2016)
Oleic acid	no	carboxyl	0.11	1.78	282.47	4.03E-06	0.128	1.00E-05	Petters (2009)

Continuation of Table S6									
Chemical	Validation Data	BAT functionality	O : C	H : C	M_{org} $(\mathrm{g\,mol^{-1}})$	BAT κ_{CCN}	AIOMFAC κ_{CCN}	Measured κ_{CCN}	Measurement Reference
Stearic acid	no	carboxyl	0.11	2.00	284.48	3.97E-06		1.00E-05	Petters (2009)
Palmitic acid	no	carboxyl	0.13	2.00	256.43	4.66E-06		1.00E-05	Petters (2009)
Myristic acid	yes	carboxyl	0.14	2.00	228.37	5.37E-06	0.053	1.00E-05	Petters (2009)
Peroxide-ether	no	hydroperoxide	0.21	2.14	246.40	3.09E-06		3.70E-03	Suda (2014)
Peroxide-ether with aldehyde	no	hydroperoxide	0.29	2.00	260.00	4.10E-06		9.20E-04	Suda (2014)
Cis-Pinonic acid	yes	carboxyl	0.30	1.60	184.24	0.054	0.106	0.005	Petters (2016)
Pinonic acid	yes	carboxyl	0.30	1.60	184.24	0.054	0.106	0.106	Raymond (2003) and Petters (2007)
Peroxide-ether with acid	no	hydroperoxide	0.36	2.00	276.40	0.000		0.020	Suda (2014)
Diperoxide-diether	no	hydroperoxide	0.43	2.14	294.40	0.000		0.011	Suda (2014)
Azelaic acid	yes	carboxyl	0.44	1.78	188.22	0.109	0.031	0.023	Petters (2009)
Homophthalic acid	yes	carboxyl	0.44	0.89	180.16	0.136	0.050	0.094	Huff Hartz (2006) and Petters (2007)
Pinic acid	no	carboxyl	0.44	1.56	187.21	0.114		0.248	Raymond (2003) and Petters (2007)
Norpinic acid	no	carboxyl	0.50	1.50	172.18	0.129	0.179	0.182	Raymond (2003) and Petters (2007)

Continuation of Table S6									
Chemical	Validation Data	BAT function-ality	O : C	H : C	M_{org} $(\mathrm{g\,mol^{-1}})$	BAT κ_{CCN}	AIOMFAC κ_{CCN}	Measured κ_{CCN}	Measurement Reference
Phthalic acid	yes	carboxyl	0.50	0.75	166.14	0.155	0.051	0.051	Huff Hartz (2006) and Petters (2007)
Pimelic acid	yes	carboxyl	0.57	1.71	160.17	0.137	0.133	0.150	Frosch (2010)
Adipic acid	no	carboxyl	0.67	1.67	146.14	0.156		0.096	Broekhuizen (2004) and Petters 2007
Polyacrylic acid	no	carboxyl	0.67	1.33	2000.00	0.017		0.054	Brooks (2004), and Petters (2009)
Polyacrylic acid	no	carboxyl	0.67	1.33	2000.00	0.017		0.051	Petters (2006, 2007)
Glutaric acid	no	carboxyl	0.80	1.80	147.13	0.157	0.133	0.106	Petters (2009)
Levoglucosan	yes	hydroxyl	0.83	1.67	162.14	0.147	0.140	0.208	Svenningsson (2006) and Petters (2007)
Maltotriose hydrate	yes	hydroxyl	0.89	1.78	504.44	0.050	0.028	0.055	Petters (2009)
Sucrose	yes	hydroxyl	0.92	1.83	342.30	0.071	0.061	0.095	Petters (2009)
alpha-Ketoglutaric acid	yes	carboxyl	1.00	1.20	146.11	0.181	0.179	0.310	Petters (2016)
Erythritol	yes	hydroxyl	1.00	2.50	122.12	0.181	0.180	0.140	Petters (2009)
Glucose	yes	hydroxyl	1.00	2.00	180.16	0.131	0.128	0.170	Petters (2009)
Maleic acid	yes	carboxyl	1.00	1.00	116.10	0.235	0.234	0.330	Petters (2016)
Succinic acid	yes	carboxyl	1.00	1.50	118.09	0.214	0.212	0.235	Petters (2009)

Continuation of Table S6

Chemical	Validation Data	BAT functionality	O : C	H : C	M_{org} $(\mathrm{g\,mol^{-1}})$	BAT κ_{CCN}	AIOMFAC κ_{CCN}	Measured κ_{CCN}	Measurement Reference
Malonic acid	no	carboxyl	1.33	1.33	104.06	0.261	0.234	0.227	Pradeep Kumar (2003) and Petters (2007)

End of Table

6 Additional Water Activity Isopleths

The non-ideal behavior of water–organic mixtures is here explored at different molecular masses of the organic, analogous to Fig. 2 of the main text. This is used to probe for any functional irregularities and was used to place bounds on realistic BAT model behavior. In Fig. S5 the isopleths for 75, 100, 150, and 200 $\mathrm{g\,mol^{-1}}$ of organic molar mass are shown. The black region in Figs. S5 – S7 represent regions of phase separation due to water activity ($a_w > 1$) and light gray those due to organic activities ($a_{org} > 1$). In the 75 $\mathrm{g\,mol^{-1}}$ case (Fig. S5a), one can start to see irregular behavior in the black phase separation region as it has a bump at O : C=0.25. The lower limit for reasonable behavior is then approximately 75 $\mathrm{g\,mol^{-1}}$ due to that irregularity – at least for $\mathrm{O:C} < 0.3$, while physically reasonable behavior is shown for higher O : C ratios. LLPS is clearly larger than the dark gray shaded areas as the 0.9 a_w contour has identical activities for two different mole fractions of water, which is indicative LLPS. Figure S6 shows the a_w-isopleths as molecular mass increases: for 300, 500, 800, and 1000 $\mathrm{g\,mol^{-1}}$. Above 500 $\mathrm{g\,mol^{-1}}$ the model is unconstrained by training data and it is at these higher molecular masses that the contours indicate artifacts due to transition effects among the distinct O : C ranges of the three BAT model domains. The dips in the a_w-contours at an O : C of about 0.1 and 0.4 in Fig.S6c & d are non-physical. Such non-physical domain transition effects are further enhanced for high molar mass compounds when the x-axis shows the mole fraction of water. To get a clearer picture of this behavior at high molecular masses, we generated isopleths graphs for 500, 800, 1000, and 2000 $\mathrm{g\,mol^{-1}}$ (Fig. S7). We changed the x-axis to a mass fraction scale to better visualize the water uptake by these large molecules. In Fig.S7b, we can start to see irregular phase separation behavior indicated by an apparent region of miscibility at $0.1 < \mathrm{O:C} < 0.15$, with phase separation at slightly higher and lower O : C. It is likely a non-physical artifact with a miscible region sandwiched between the black regions; it should very likely be one contiguous phase separation region. This irregular behavior then continues to expand as the molecular weight increases in Fig.S7c & d. However, we emphasize here that the gray areas only show the minimum extent of an LLPS region, while a liquid–liquid equilibrium computation (as done with VBS + BAT) needs to be done to determine the thermodynamically favoured parameter space exhibiting LLPS. If one is interested in phase separation predictions and BAT calculations for organics of $\mathrm{O:C} < 0.45$, then the BAT model is limited to the molar mass range below 750 $\mathrm{g\,mol^{-1}}$. If one is only interested in the O : C region above 0.5, then the BAT model should be applicable, with reasonable behavior exhibited up to at least 2000 $\mathrm{g\,mol^{-1}}$.

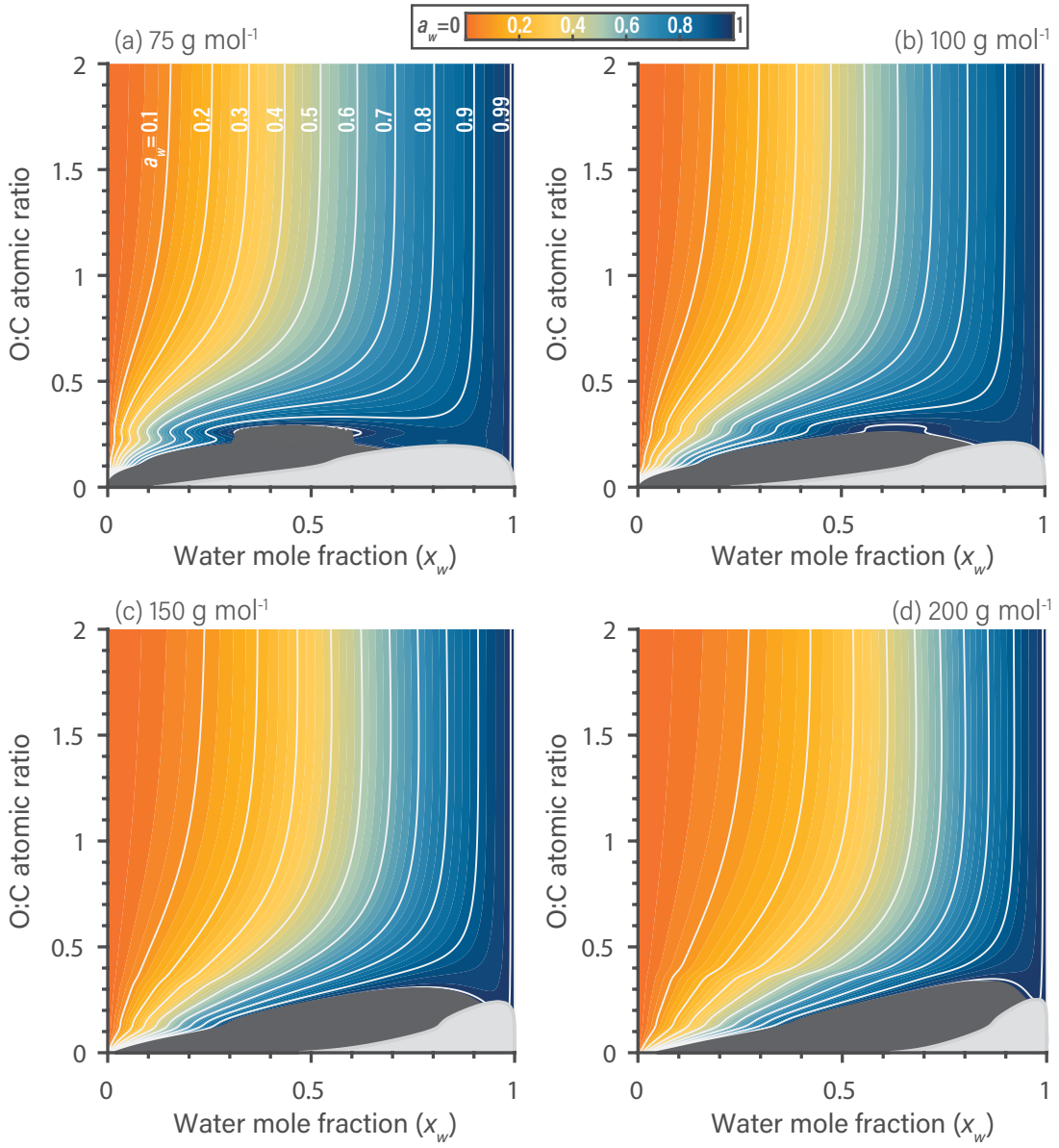

Figure S5. Predicted water activity contours generated by the BAT model for binary aqueous mixtures of generic organic compounds of constant molar mass yet variable O : C at $T = 298.15$ K. The contours link water mole fraction and the organic O : C to the resulting water activity in a binary water–organic mixture. The combined shaded regions in dark ($a_w > 1$) and light gray ($a_{org} > 1$) represent the minimum extent of liquid–liquid phase separation for a certain O : C. The bumps in the contours at O : C of 0.1 and 0.3 stem from the transitions between the BAT model's low-, medium-, and high-O : C parameterization domains. The M_{org} used is as follows: (a) 75 $\mathrm{g\,mol^{-1}}$, (b) 100 $\mathrm{g\,mol^{-1}}$, (c) 150 $\mathrm{g\,mol^{-1}}$, and (d) 200 $\mathrm{g\,mol^{-1}}$.

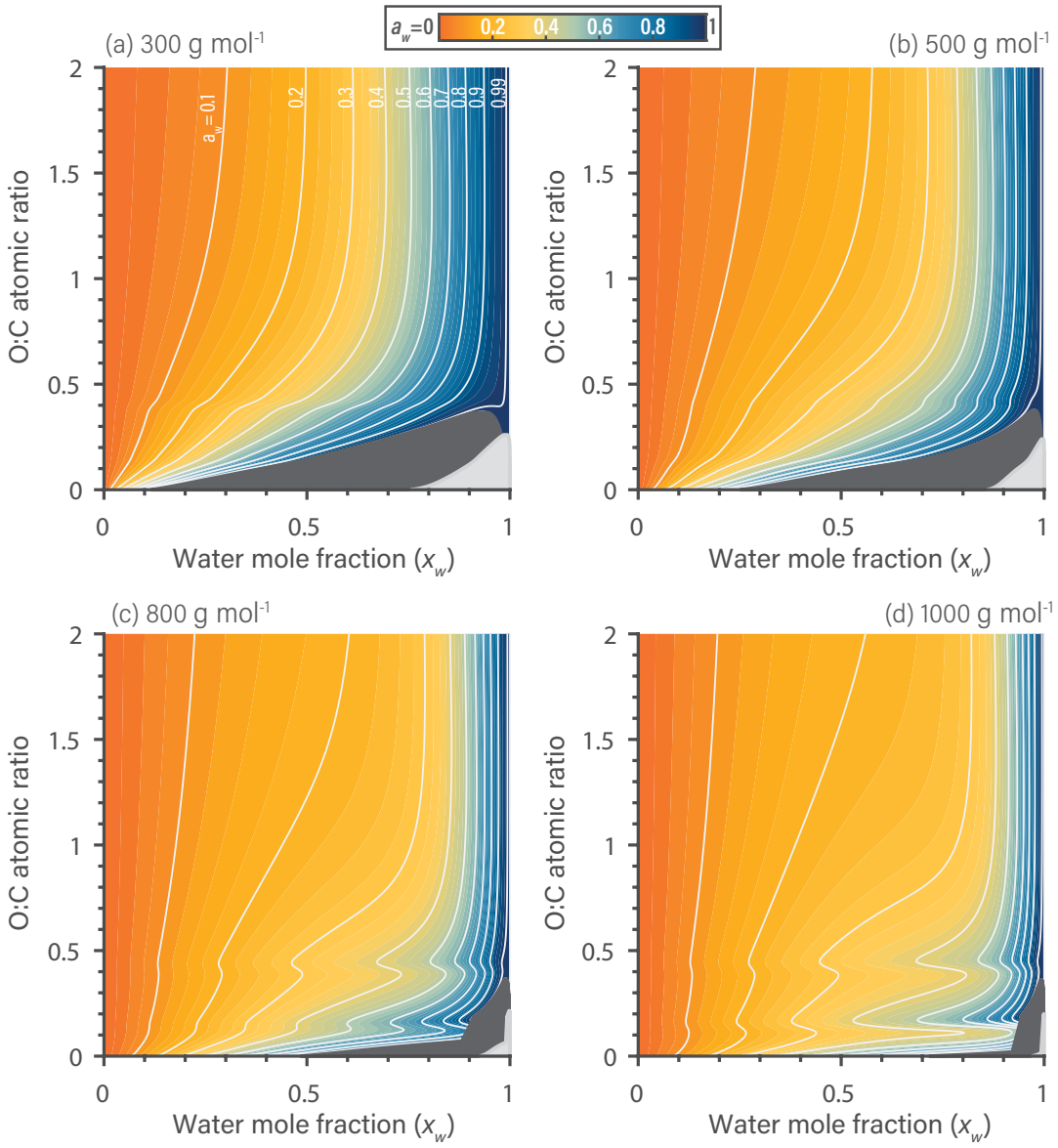

Figure S6. Predicted water activity contours generated by the BAT model for binary aqueous mixtures of generic organic compounds of constant molar mass yet variable O : C at $T = 298.15$ K. The contours link water mole fraction and the organic O : C to the resulting water activity in a binary water–organic mixture. The combined shaded regions in dark ($a_w > 1$) and light gray ($a_{org} > 1$) represent the minimum extent of liquid–liquid phase separation for a certain O : C. The bumps in the contours at O : C of 0.1 and 0.45 stem from the transitions between the BAT model's low-, medium-, and high-O : C parameterization domains. The M_{org} used is as follows: (a) 300 g mol^{-1}, (b) 500 g mol^{-1}, (c) 800 g mol^{-1}, and (d) 1000 g mol^{-1}

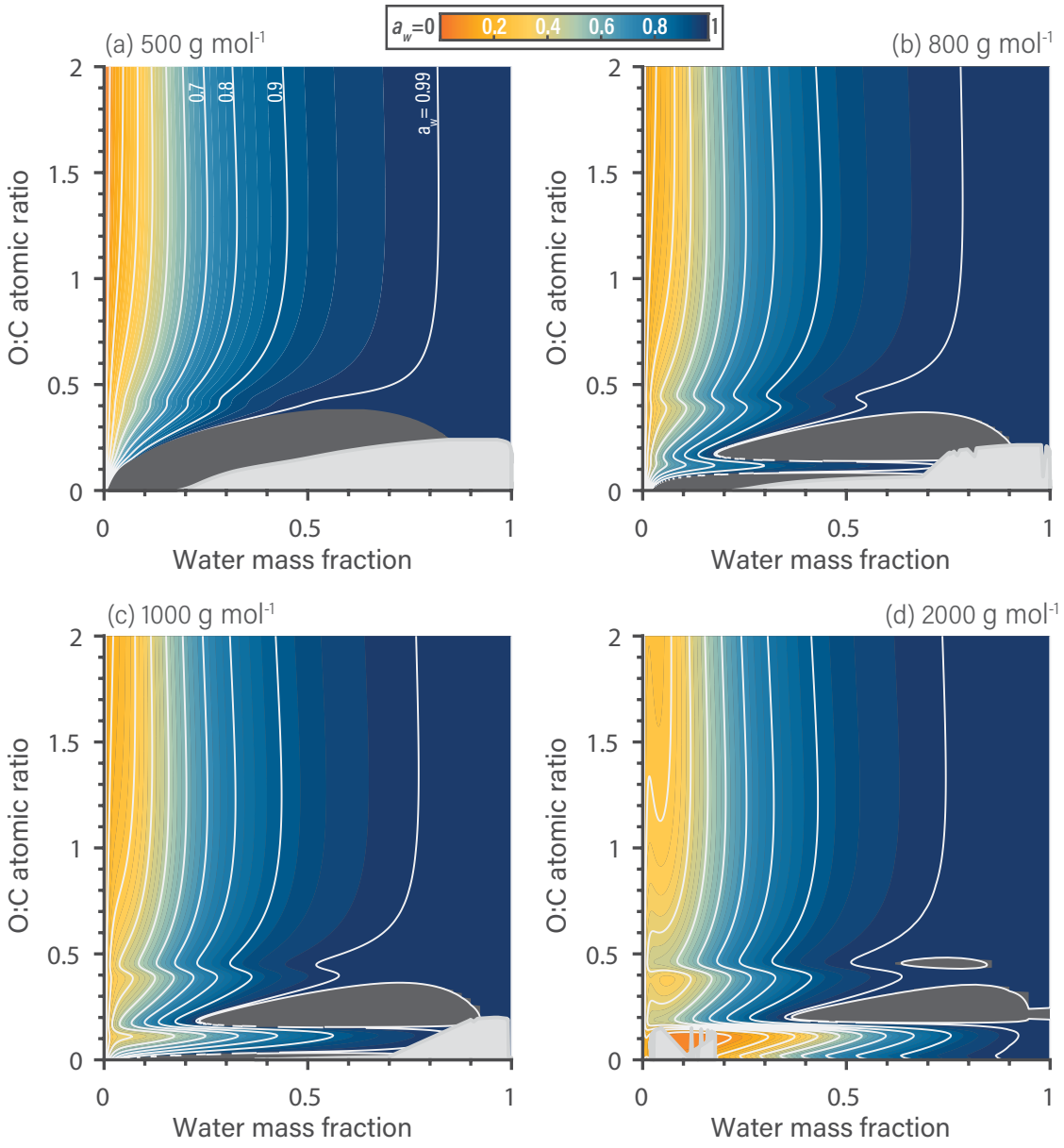

Figure S7. Predicted water activity contours generated by the BAT model for binary aqueous mixtures of generic organic compounds of constant molar mass yet variable O : C at $T = 298.15$ K. **Note the change to a mass fraction scale.** The contours link water mass fraction and the organic O : C to the resulting water activity in a binary water–organic mixture. The combined shaded regions in dark ($a_w > 1$) and light gray ($a_{org} > 1$) represent the minimum extent of liquid–liquid phase separation for a certain O : C. The bumps in the contours at O : C of 0.1 and 0.45 stem from the transitions between the BAT model's low-, medium-, and high-O : C parameterization domains. The M_{org} used is as follows: (a) 500 g mol^{-1}, (b) 800 g mol^{-1}, (c) 1000 g mol^{-1}, and (d) 2000 g mol^{-1}

7 SOA Mixtures

The model comparison focuses on the predictions of bulk liquid aerosol mass concentration, and we used the AIOMFAC-based equilibrium gas–particle partitioning predictions as a benchmark. The AIOMFAC-equil. calculations include consideration of liquid–liquid phase separation and consider relatively high-fidelity input, as the AIOMFAC model uses functional group information and accounts for non-ideal interactions among all species. In contrast, the VBS + BAT approach only includes non-ideal water $\leftrightarrow$ organic interactions (implicitly assuming ideal organic $\leftrightarrow$ organic mixing) and rather limited molecular structure information ($\mathrm{O:C}$ and M_{org}). The full extent of the percentage difference in organic aerosol mass between the VBS + BAT approach and AIOMFAC-equil. is shown in Fig. S8.

For our simulated aerosol systems, we use surrogate systems representing α-pinene SOA (Table S7) and isoprene SOA (Table S8) products based on predictions from the Master Chemical Mechanism, as was detailed in Zuend and Seinfeld (2012) and Chen et al. (2011), respectively. The α-pinene SOA system used here contains 10 organic species as surrogates of the SOA, and the isoprene SOA system is comprised of 21 organic surrogate species. The input $\mathrm{O:C}$ and M_{org} used for BAT are listed in Tables S7 & S8 and the molecular functionality translations to OH-equivalents (done internally in the model) are listed in square brackets.

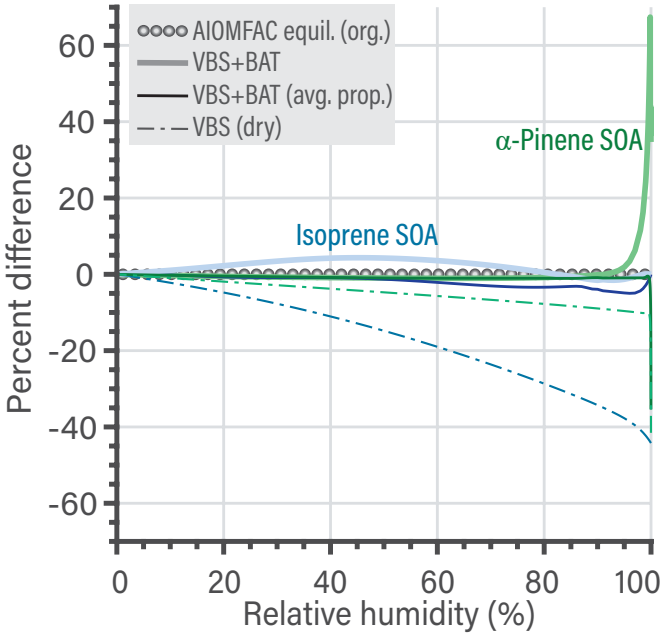

Figure S8. Percent difference in organic aerosol mass between the VBS + BAT approach and AIOMFAC-equil. as a function of equilibrium relative humidity for a bulk solution ($= a_w$) at 298.15 K. Simulations for isoprene SOA are shown in blue and those for α-pinene SOA in green. The benchmark AIOMFAC equilibrium predictions are shown for the salt-free cases (circles). The thick curves show the VBS + BAT prediction with different organic components, while the thin curve shows a simulation assuming an average molecule calculated from the dry mass, i.e., average O : C, H : C, M_{org}, and we kept the individual molecule's effective C^{sat}_{dry}. The thin dashed line shows the percent difference in the standard VBS simulation with no water uptake (dry).

Table S7: Properties of the α-pinene SOA organic mixture used. The brackets denote the BAT model's internal molecular functionality translation.

Start of Table S7							
MCM Name	SMILES	BAT functionality	O : C [OH eqv.]	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$) [OH eqv.]	$C^{g+\Sigma_\pi}$ ($\mu\mathrm{g\,m^{-3}}$)	eff. C^{sat}_{dry} ($\mu\mathrm{g\,m^{-3}}$)
C107OOH	O=CCC1CC(OO)(C(=O)C)C1(C)C	hydroperoxideSOA	0.40 [0.22]	1.60	200.17 [164.22]	8.7918E+00	5.7429E+03
C97OOH	OCC1CC(OO)(C(=O)C)C1(C)C	hydroperoxideSOA	0.44 [0.24]	1.78	188.17 [152.78]	3.9840E+00	3.2741E+02

Continuation of Table S7							
MCM Name	SMILES	BAT functionality	O : C [OH eqv.]	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$) [OH eqv.]	$C^{g+\Sigma_\pi}$ ($\mu\mathrm{g\,m^{-3}}$)	eff. C^{sat}_{dry} ($\mu\mathrm{g\,m^{-3}}$)
C108OOH	O=CCC(CC(=O)C(=O)C)C(C)(C)OO	hydroperoxideSOA	0.50 [0.27]	1.60	216.13 [216.13]	1.1344E+00	1.6671E+02
PINIC	OC(=O)CC1CC(C(=O)C)C1(C)C	carboxyl	0.44 [0.44]	1.56	186.17 [186.17]	6.2815E-01	1.4953E+01
C921OOH	OCC(=O)C1(OO)CC(CO)C1(C)C	hydroperoxideSOA	0.56 [0.30]	1.78	204.18 [168.09]	9.1858E-01	2.1280E+00
C812OOH	OCC1CC(OO)(C(=O)O)C1(C)C	hydroperoxideSOA	0.86 [0.46]	1.75	195.17 [159.44]	7.6636E-01	7.1911E-01
C811OH	OCC1CC(C(=O)O)C1(C)C	hydroperoxideSOA	0.38 [0.20]	1.75	158.17 [124.84]	3.9949E-01	1.1569E+03
C813OOH	OCC(CC(=O)C(=O)O)C(C)(C)OO	hydroperoxideSOA	0.75 [0.40]	1.75	206.14 [169.98]	3.1319E-01	3.0180E-02
ALDOL-dimer	CC(=O)C(=O)CC(C(C=O)=CCC1CC(C(O)=O)C1(C)C)C(C)(C)OO	hydroperoxideSOA	0.37 [0.20]	1.47	368.30 [335.21]	4.0696E+00	2.7866E-06
ESTER-dimer	CC1(C)C(CC1C(O)=O)CC(=O)OCC(=O)C2CC(CC(O)=O)C2(C)C	ester	0.37 [0.12]	1.56	368.31 [289.50]	1.0174E+00	3.6370E-06
End of Table							

Table S8: Properties of the isoprene SOA organic mixture used. The brackets denote the BAT model's internal molecular functionality translation.

Start of Table S8							
MCM Name	SMILES	BAT functionality	O : C [OH eqv.]	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$) [OH eqv.]	$C^{g+\Sigma_\pi}$ ($\mu\mathrm{g\,m^{-3}}$)	eff. C^{sat}_{dry} ($\mu\mathrm{g\,m^{-3}}$)

Continuation of Table S8							
MCM Name	SMILES	BAT functionality	O : C [OH eqv.]	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$) [OH eqv.]	$C^{g+\Sigma_\pi}$ ($\mathrm{\mu g\,m^{-3}}$)	eff. C^{sat}_{dry} ($\mathrm{\mu g\,m^{-3}}$)
IEB1OOH	OCC(O)C(C)(OO)C=O	hydroperoxideSOA	1.00 [0.54]	2.00	150.11 [117.51]	3.2124E+00	5.0688E+01
IEB2OOH	OOC(C=O)C(C)(O)CO	hydroperoxideSOA	1.00 [0.54]	2.40	150.11 [117.51]	2.4919E-01	2.3180E+02
C59OOH	OCC(=O)C(C)(CO)OO	hydroperoxideSOA	1.00 [0.54]	2.00	150.09 [117.50]	4.2176E+00	2.2954E+01
IEC1OOH	OCC(=O)C(C)(CO)OO	hydroperoxideSOA	1.00 [0.54]	2.00	150.09 [117.50]	1.4709E+00	2.2954E+01
C58OOH	O=CC(O)C(C)(CO)OO	hydroperoxideSOA	1.00 [0.54]	2.00	150.11 [117.51]	3.3475E-01	5.0688E+01
IEPOXA	CC(O)(CO)C1CO1	hydroxyl	0.60	2.00	118.13	8.6354E-11	3.5120E+13
C57OOH	OCC(O)C(C)(OO)C=O	hydroperoxideSOA	1.00 [0.54]	2.00	150.11 [117.51]	2.7170E-01	5.0688E+01
IEPOXC	CC1(CO1)C(O)CO	hydroxyl	0.60 [0.60]	2.00	118.13 [118.13]	2.7879E-09	5.2036E+04
HIEB1OOH	OCC(O)C(CO)(OO)C=O	hydroperoxideSOA	1.20 [.64]	2.00	166.11 [132.13]	2.8903E-01	1.0370E-01
INDOOH	OCC(ON(=O)=O)C(C)(CO)OO	hydroperoxideSOA	1.40 [0.75]	2.20	197.14 [161.32]	2.5037E-01	4.5117E-01
IEACO3H	CC(O)(C1CO1)C(=O)OO	hydroperoxideSOA	1.00 [0.54]	1.60	148.10 [115.69]	5.3463E-08	5.6321E+04
C525OOH	OCC(=O)C(CO)(CO)OO	hydroperoxideSOA	1.20 [0.64]	2.00	166.09 [132.12]	2.1592E-01	3.9838E-02
HIEB2OOH	OOC(C=O)C(O)(CO)CO	hydroperoxideSOA	1.20 [0.64]	2.00	166.11 [132.13]	1.4203E-01	7.0484E-01
IEC2OOH	OCC(=O)C(C)(OO)C=O	hydroperoxideSOA	1.00 [0.54]	1.60	148.06 [115.66]	2.0876E-06	4.2944E+03
INAOOH	OCC(C)(OO)C(O)CON(=O)=O	hydroperoxideSOA	1.40 [0.75]	2.20	197.14 [161.32]	1.3898E-01	1.7351E+00

Continuation of Table S8							
MCM Name	SMILES	BAT functionality	O : C [OH eqv.]	H : C	M_{org} ($\mathrm{g\,mol^{-1}}$) [OH eqv.]	$C^{g+\Sigma_\pi}$ ($\mu\mathrm{g\,m^{-3}}$)	eff. C^{sat}_{dry} ($\mu\mathrm{g\,m^{-3}}$)
C510OOH	O=CC(O)C(C)(OO)CON(=O)=O	hydroperoxideSOA	1.40 [0.75]	1.8	195.10 [159.38]	4.1752E-03	2.6990E+02
INB1OOH	OCC(OO)C(C)(CO)ON(=O)=O	hydroperoxideSOA	1.40 [0.75]	2.20	197.14 [161.32]	7.1561E-02	4.2126E-01
IECCO3H	CC1(CO1)C(O)C(=O)OO	hydroperoxideSOA	1.00 [0.54]	1.60	148.11 [115.71]	7.5983E-07	1.8033E+04
INCOOH	OCC(OO)C(C)(O)CON(=O)=O	hydroperoxideSOA	1.40 [0.75]	2.20	197.14 [161.32]	3.0754E-02	7.3141E+00
INB2OOH	OOCC(O)C(C)(CO)ON(=O)=O	hydroperoxideSOA	1.40 [0.75]	2.20	197.14 [161.32]	3.4893E-02	1.4651E+00
2-Methyltetrol-dimer	CC(O)(CO)C(O)COC(C)(CO)C(O)CO	hydroxyl	0.70 [0.70]	2.30	254.28 [254.28]	7.2215E+00	2.5788E-06
End of Table							

References

Broekhuizen, K., Kumar, P. P., and Abbatt, J. P.: Partially soluble organics as cloud condensation nuclei: Role of trace soluble and surface active species, Geophys. Res. Lett., 31, 1–5, https://doi.org/10.1029/2003GL018203, 2004.

Brooks, S. D., DeMott, P. J., and Kreidenweis, S. M.: Water uptake by particles containing humic materials and mixtures of humic materials with ammonium sulfate, Atmos. Environ., 38, 1859–1868, https://doi.org/10.1016/j.atmosenv.2004.01.009, 2004.

Chen, Q., Liu, Y., Donahue, N. M., Shilling, J. E., and Martin, S. T.: Particle-phase chemistry of secondary organic material: Modeled compared to measured O:C and H:C Elemental ratios provide constraints, Environ. Sci. Technol., 45, 4763–4770, https://doi.org/10.1021/es104398s, 2011.

Frosch, M., Zardini, A. A., Platt, S. M., Müller, L., Reinnig, M.-C., Hoffmann, T., and Bilde, M.: Thermodynamic properties and cloud droplet activation of a series of oxo-acids, Atmos. Chem. Phys., 10, 5873–5890, https://doi.org/10.5194/acp-10-5873-2010, 2010.

Girolami, G. S.: A Simple "Back of the Envelope" Method for Estimating the Densities and Molecular Volumes of Liquids and Solids, J. Chem. Educ., 71, 962–964, https://doi.org/10.1021/ed071p962, 1994.

Huff Hartz, K. E., Tischuk, J. E., Chan, M. N., Chan, C. K., Donahue, N. M., and Pandis, S. N.: Cloud condensation nuclei activation of limited solubility organic aerosol, Atmos. Environ., 40, 605–617, https://doi.org/10.1016/j.atmosenv.2005.09.076, 2006.

Petters, M. D., Kreidenweis, S. M., Snider, J. R., Koehler, K. A., Wang, Q., Prenni, A. J., and Demott, P. J.: Cloud droplet activation of polymerized organic aerosol, Tellus, Ser. B Chem. Phys. Meteorol., 58, 196–205, https://doi.org/10.1111/j.1600-0889.2006.00181.x, 2006.

Petters, M. D., Kreidenweis, S. M., Prenni, A. J., Sullivan, R. C., Carrico, C. M., Koehler, K. A., and Ziemann, P. J.: Role of molecular size in cloud droplet activation, Geophys. Res. Lett., 36, 1–5, https://doi.org/10.1029/2009GL040131, 2009.

Petters, M. D., Kreidenweis, S. M., and Ziemann, P. J.: Prediction of cloud condensation nuclei activity for organic compounds using functional group contribution methods, Geosci. Model Dev., 9, 111–124, https://doi.org/10.5194/gmd-9-111-2016, 2016.

Pradeep Kumar, P., Broekhuizen, K., and Abbatt, J. P.: Organic acids as cloud condensation nuclei: Laboratory studies of highly soluble and insoluble species, Atmos. Chem. Phys., 3, 509–520, https://doi.org/10.5194/acp-3-509-2003, 2003.

Raymond, T. M.: Formation of cloud droplets by multicomponent organic particles, J. Geophys. Res., 108, 4469–4477, https://doi.org/10.1029/2003JD003503, 2003.

Suda, S. R., Petters, M. D., Yeh, G. K., Strollo, C., Matsunaga, A., Faulhaber, A., Ziemann, P. J., Prenni, A. J., Carrico, C. M., Sullivan, R. C., and Kreidenweis, S. M.: Influence of Functional Groups on Organic Aerosol Cloud Condensation Nucleus Activity, Environ. Sci. Technol., 48, 10 182–10 190, https://doi.org/10.1021/es502147y, 2014.

Svenningsson, B., Rissler, J., Swietlicki, E., Mircea, M., Bilde, M., Facchini, M. C., Decesari, S., Fuzzi, S., Zhou, J., Mønster, J., and Rosenørn, T.: Hygroscopic growth and critical supersaturations for mixed aerosol particles of inorganic and organic compounds of atmospheric relevance, Atmos. Chem. Phys., 6, 1937–1952, https://doi.org/10.5194/acp-6-1937-2006, 2006.

Zuend, A. and Seinfeld, J. H.: Modeling the gas-particle partitioning of secondary organic aerosol: The importance of liquid-liquid phase separation, Atmos. Chem. Phys., 12, 3857–3882, https://doi.org/10.5194/acp-12-3857-2012, 2012.