1 Anonymous Referee #1

Gorkowski et al. developed a very useful model BAT that can treat the non-ideal mixing of organics and water and can predict the liquid-liquid phase separation, which is very important in SOA partitioning. The BAT model uses the measurable organic aerosol properties (oxidation state, molar mass and vapor pressure) as inputs and the simulated results agree with a comprehensive thermodynamic model AIOMFAC. The BAT model is successfully coupled with the VBS model predicting the gas-particle partitioning. The topic of this study is timely and highly relevant in improvement of thermodynamic aerosol treatment in chemical transport models. I recommend this manuscript for publication after the following comments can be addressed.

Authors Response: We thank the reviewer for her/his appreciation of this work and support of the manuscript. We have added clarifying text about the functional group translation approach and possible extensions to it. Below are our responses to specific comments, with reviewer comments in purple, our responses in black text, and changes to the manuscript showing removed text in red with strikethroughs and added text in blue with underlines. The page numbers listed after "Manuscript Revisions on Page" refer to the revised manuscript.

1.1 Major comments

My major concern goes to the method of Functional Group Translation: P12, Line 10-11: Can the "functional group translation" also treat the nitrogen or sulfur-bearing functional groups?

Authors Response: This version of the BAT model does not include a parameter set to account explicitly for sulfur or nitrogen moieties. We have expanded on the functional group translation discussion in the main text to point out possible extensions to it. We also changed the phrase to 'molecule functionality translation' since there could be confusion with a traditional group-contribution approach (like in UNIFAC / AIOMFAC) accounting for individual functional group effects rather than a whole molecule's effect based on a predominant functionality classification. Our method is not an individual functional group translation, though it may be possible to develop one, right now a whole molecule is assigned a fixed set of translation coefficients to translate the BAT parameterization when the predominant functionalities of a molecule (or class of molecules when averaging) is known in a system and when it is different from hydroxyl-dominated molecules.

Manuscript Revisions on Page 14–15: *We reiterate that the BAT model is describing the whole molecule, and so these translations are not for the individual functional groups ~~on a multifunctional molecule~~ . This method is different from the group contribution approach taken by UNIFAC and AIOMFAC, as here the whole molecule is assigned one effective functionality. For multifunctional molecules, a distinct multifunctional translation ~~must~~ may be derived, like we did for the SOA oxidation products (see Fig. 4b). ~~If that is not possible~~ This can be done by using AIOMFAC to generate training data for multifunctional molecules that are representative of VOC oxidation products. The molecular translation coefficients are then fitted using the generated training database. If this fitting of the translation coefficients is not practical, then the ~~most dominant and most representative oxygen functionality~~ predominant or most representative oxygen-bearing functionality on the molecule should be chosen ~~.~~ for an approximate molecule functionality translation. Extensions to include organic nitrate and sulfate functionalities will be a topic of future development. In principle, additional molecular functionality translations for each combination of molecular functionalities could be developed, which would be practical if the number of permutations is small. If the number*

of combinatorial permutations of molecular functionalities is large, then that development direction would lead to increased complexity, which is not the goal of the BAT model. We will explore different weighting and scaling methods of the translations coefficients based on $N : C$ *and* $S : C$ *elemental ratios to retain the reduced-complexity approach. If accurate activity coefficient predictions of a known set of multifunctional molecules are desired and the molecular structures are known, then the use of AIOMFAC or a system specific model instead of BAT is recommended.*

P14, Line 8-10: I suggest adding a more detailed description to explain how to do "a distinct multifunctional translation". How the functional group translation is calculated for C97OOH in Fig.4(b)? The translated O:C ratio and molar mass can be added in Tables S5 and S6 in the supplement.

Authors Response: See the response to the comment above and related changes to the manuscript. As suggested, we have also added translated values to the SI tables.

P14, Line 22-25: Is Fig.4a based on the carboxy-based, ketone-based, etc parameterizations? The shaded grey area and the pink line in Fig. 4a are not explained in the main text. Please help me understand Fig.4a.

Authors Response: Yes, Fig. 4a shows the effect of BAT parameterizations with regard to the limit-of-miscibility lines when based on the indicated functional groups rather than hydroxyl. The dotted pink line is the result of applying the multifunctional hydroperoxide translation, which is used in Fig. 4b. This is stated in the figure caption and we now mention it in the text too. The grey area is the error in the $O : C$ prediction, this is explained in the figure caption, and we also added a clarifying note in the main text for this.

Manuscript Revisions on Page 15: *We use* ~~this translation~~ *such translations to plot the* ~~limit of miscibility~~ *limit-of-miscibility lines for all of the fitted functional group types considered (Fig. 4a). The dotted pink line is from the multifunctional hydroperoxide translation and the gold line is from the PEG translation, both have example translations shown in Fig. 4b and 4c respectively. The uncertainty range in the* $O : C$ *prediction of a limit of miscibility is also shown in Fig. 4a as a shaded gray region. These miscibility limit lines represent the same process (phase separation limit), but for different functional groups, so it is informative to compare their relative positions* ~~.~~ *in Fig. 4a.*

Minor comments:

(1) P5, Line 21: It is not proper to describe Eq. (6) as the effective volatility of "all species". It is still the effective volatility of the compound j but includes water and inorganics in the absorbing phase.

Authors Response: We meant to state that it applies to all species; phrasing amended.

Manuscript Revisions on Page 5: *The effective* ~~volatility of all~~ *saturation concentration of each species, including water and other inorganic constituents in liquid phase π, is defined by Eq. (6).*

(2) P9, Line 9: Could the authors explain more how you get the scaling factor in the form of [s1(1+O:C)s2]? From Section 3.2 it seems s1 and s2 are fitted by the training dataset generated by the AIOMFAC model, instead of experimental data as you wrote here on Line 9.

Authors Response: You are correct, it is fitted with the AIOMFAC training dataset.

Manuscript Revisions on Page 9: *In Eq. (16), ρ_{org} and ρ_w are the liquid-state densities of the organic component and water, respectively, while s_1 and s_2 are two scaling parameters determined ~~from a model fit to experimental~~ during the model fitting to training data.*

(3) P10, Line 31: The authors wrote "the light green domain starts at $\sim 20\%$ of the O:C ratio reached at the miscibility limit and covers up to the blue domain", but from Table S1, it seems the light green domains starts from O:C of 0.05 and covers up to the O:C of 10% of the miscibility line? In the excel file, the mid O:C region is "$0.05 < O:C < 0.1+$ miscibility line", which is different from Table S1 ($0:05 < O:C < 0.1$ miscibility line).

Authors Response: The excel file $O:C$ bounds are correct, the SI has been revised to match this.

Table S1 range is the data used in the fitting of the coefficients but that range does not directly map to the coefficients used in a BAT model calculation. A weighted averaged of the coefficients (SI section 2.3) are used in the BAT calculation; an example graph has been added to the SI. The text has been clarified accordingly.

Manuscript Revisions on Page 11: *The blue domain includes components that have no miscibility limit with water. The light green domain starts at ~~$\sim 20\%$~~ $\sim 30\%$ of the $O:C$ ratio reached at the miscibility limit and covers up to the blue domain.* $\cdots$

~~Parameter optimization was carried out separately for each of the three domains, resulting in three distinct sets of BAT model parameters.~~ *These domains represent the three regions where each set of optimized parameters dominates. Parameter optimization for each sets of coefficients was carried out on a wider and overlapping $O:C$ range than shown in Fig. 2a.* $\cdots$

An example of the sigmoidal transition function is shown in the SI, Fig. S1.

(4) P11, Line 8-10: Could the authors explain in a more detailed way how the equation (S14) is derived to calculate the limit of miscibility line? How you determined the O:C range of 0.05 to 0.45?

Authors Response: We added text to the SI (page 3, Sect. 2.2) to explain our approach for this.

SI Revisions on Page 3: *The limit of miscibility line is determined from an initial BAT model ~~fitting~~ 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 ~~defining~~ were used to fit the limit of miscibility line, ϑ_{ML}, as a function of organic molar mass, ~~was determined as~~*

(5) P11, Line 29: The sentence is correct but the (aw > xw) confuses me as from Fig.3, for the higher O:C region, the predicted aw is smaller than xw.

Authors Response: Thanks for spotting this, you are correct. The sentence has been revised.

Manuscript Revisions on Page 12: *Moving up towards higher $O:C$, there is a transition to rather hydrophilic behavior and the water uptake at given equilibrium RH is predicted to become higher than that of an ideal mixture (~~$a_w > x_w$~~ $a_w < x_w$).*

(6) P12, Line 6-7: I couldn't see this result from Fig.3 and I don't quite understand the grey areas in Fig.3. Could the authors help explain it?

Authors Response: The increase with M_{org} can also be seen in Fig. 4a (hydroxyl curve) and from the additional isopleth plots we added to the SI with this revised version of the manuscript.

The gray areas mark LLPS regions due to either activity (organic or water) being greater than 1, e.g. a pure organic phase would have $a_{org} = 1$, which would therefore favor phase separation whenever the (forced) mixing leads to activities of one or several components exceeding 1. LLPS also occurs where an identical activity (either a_w or a_{org}) is predicted for two different mole fractions of water. In this figure, the gray shading indicates that an initial binary mixture within the gray area would be unstable and undergo LLPS, leading to two phases of distinct compositions.

Manuscript Revisions on Page 13: *In a binary mixture, LLPS is also clearly indicated anywhere a component activity is (predicted) to be greater than* 1.0 *when assuming a single liquid phase in the calculation* ~~.~~ *(gray areas in Fig. 3). These gray areas mark initial compositions that would be unstable and quickly lead to separation into two phases of distinct water mole fractions; in the case of Fig. 3 with the final phase compositions given by the two intersection points of a line of constant* $O:C$ *(of compound in question) and the water activity contour at the edge of the phase separation area. Additional isopleths at different organic molar masses (75 to 2000* $\mathrm{g\,mol^{-1}}$*) are shown in the SI Sect. 6. Based on BAT predictions, in comparison to the case shown in Fig. 3, this phase separation region moves to higher* $O:C$ *as the molar mass of the organics increases and to lower* $O:C$ *as molar mass decreases.*

(7) P14, Line 11: It is better to describe the Fig.4 from Fig.4(a) to (c).

Authors Response: We went with Fig.4b & c first as they are direct examples, which follow clearly from the introduction of the translation methodology in the main text. We don't think it is necessary for the figure description to be chronological with the main text. We have considered changing the order of the Fig 4. (b & c graphs on the left then a), but we think it is more aesthetically pleasing the way it is.

(8) Figure 6: Should ξj be ξjguess in the output of the VBS neural network? I also suggest add aw,sep in the program outline.

Authors Response: We added $a_{w,sep}$ and we changed ξj to ξjguess.

1.2 Technical corrections

(1) P2, Line 31: "remains" should be "remain". **Authors Response:** Changed.

(2) P10, Line15: should be organic $\leftrightarrow$ organic interactions. The latter "organic" is missing.

Authors Response: Added.

(3) P31, Line1: There are two "the" at the beginning of the sentence.

Authors Response: fixed.

2 Anonymous Referee #2

Gorkowski et al. present a modeling approach to predict the water content, CCN activity, liquid-liquid phase separation, and gas-particle partitioning of single component and mixed organic aerosol. The focus of the work is to produce reduced complexity models that have fast runtime while preserving the fidelity of the predictions. This is achieved by training the reduced complexity model using more computationally expensive modeling framework. This manuscript is an ambitious attempt to contribute to the efficiency of modeling a wide range of organic aerosol processes. Constructing such comprehensive and fast models is technically demanding and the authors should be complimented for their often clever approaches. For example, conceiving and finding a suitable set of fitting coefficients that represent OA through Eq. (18) and (19) is impressive. A selected set of validations is presented, and these validations appear to demonstrate that the reduced complexity models are adequate. However, I do have concerns about the stability and validation of the model. A detailed formal evaluation of the BAT and NN model that is independent of training data is needed. Furthermore, more systematic validation of the model predictions against experimental data is needed, especially against single component CCN data. I anticipate that the paper will be acceptable for publication if formal, systematic, and independent validation is included.

Authors Response: We thank the reviewer for her/his positive comments and the concerns about model validation. Our description did indeed not include finer details about the splits of the database into training/validation/testing data in the context of fitting the neural networks and of BAT. Although, such procedures were followed during development; we have improved the description in the manuscript in this regard. In the revised manuscript version, we have added independent validation data for fitting the BAT coefficients and explored the stability of the BAT model through additional plots added to the SI. Additional text was added to describe the training of the neural networks using BAT-generated random data which was then separated into training (70 %), validation (15 %), and testing (15 %) data sets. Below are our responses to specific comments, with reviewer comments in purple, our responses in black text, and changes to the manuscript showing removed text in red with strikethroughs and added text in blue with underlines. The page numbers listed after "Manuscript Revisions on Page XX:" refer to the revised manuscript.

2.1 Major comments

Both the BAT model and the NN model are trained. Figure 2a shows the training points for the BAT model. The standard approach in machine learning is to have a training set for which the model is optimized, and a validation set for compounds that the model has not been tuned to. This does not appear to have been done and one might seriously question the fidelity of the model outside the training set. Showing activity for citric acid is insufficient. I recommend that the authors test the model against 100 (or so) compounds that were not used in the optimization and show a scatter plot of AIOMFAC vs. BAT for activity coefficient at various RH, mole-fraction of the predicted miscibility gap, water activity of the miscibility gap, and predicted kappa CCN. Only through independent tests and systematic characterization of the error can one be confident that the BAT coefficients truly represent AIOMFAC. A similar case is to be made for the NN training. Systematic validation against with non-training data needs to be presented.

Authors Response:

The reliability and validation of both BAT and NN models were assessed as outlined more specifically below. We note here

that there is a distinction between the training of an unknown multivariate function, as in the case of a NN, requiring a machine learning approach to determine the functional form (i.e. number of neurons and layers) and fit associated activation function parameters – and that of the BAT model, for which a thermodynamic model was prescribed with fixed coefficient functions (after initial tests), which then becomes a classical parameter optimization problem.

- **Neural Networks:** We have added a more detailed description of the training of the neural networks, as we did use the standard practice of training, validation, and test data sets. Also note that in the case of the NN, these are random data sets generated by the function the NN is trying to invert and not measurements or AIOMFAC-generated data. Hence, a large number of data points ($\sim$ 10 million) were generated for the NN training and validation.

    Manuscript Revisions on Page 20: *To fit the neuron activation functions, we generate a random data set of $O : C_j$, $M_{org,j}$, $x_{org,j}$, and a_w using the BAT model. The data corresponding to systems with a miscibility gap are parsed into two separate categories to train a separate NN. We generated a database of 9.8×10^6 data points for miscible organics and 4.6×10^5 data points for phase separated systems. Each database was then split into training data (70 %), validation data (15 %), and test data (15 %), which was used to train the BAT-NN. Our NN inputs are $O : C_j$, $M_{org,j}$, and a_w with $x_{org,j}$ as the target output.*

    And: *We tested different NN input combinations and settled on using $C_j^{g+\Sigma_\pi}$, $O : C_j$, $M_{org,j}$, BAT-derived water mass fraction ($w_{w,j}$) and a_w associated with organic component j. Using the VBS + BAT equilibrium solver, we generated a random database of 13,000 data points split into training data (70 %), validation data (15 %), and test data (15 %). This generated database was then used for the training of the NN. The NN output target is the vector of partitioning coefficients, which is subsequently used as the initial guess for solving the coupled VBS + BAT system of non-linear equations.*

- **BAT model:** Given that the BAT model is a multivariate function, a validation data set for it is also a good suggestion. We have added the following clarifications.

    Manuscript Revisions on Page 10-11: *We generated a database of 37 known organic chemical structures and 123 artificial, yet possible chemical structures. There were an additional 16 organic chemicals used for a validation database (SI Table S6), and therefore not included in the fitting of the model.*
    . . .
    For each structure there are an additional 40 data points at varying mole fractions, which means the training database has 6400 points and the validation database has 640 points.
    . . . *Page 11* . . .
    Generally, the BAT model showed good agreement to the training database with a root mean squared error (RMSE) in a_w of 0.058 (5.8 % RH) and in organic activity (a_{org}) of 0.090. The validation database showed a similar agreement with a RMSE in a_w of 0.066 and in a_{org} of 0.096 (details in SI Sect. 5). The BAT model is valid for organic molecules within the following domain: $0 \leq O : C \leq 2$ and $75 \leq M_{org} \leq 500 \, \text{g mol}^{-1}$ with realistic behavior up to $750 \, \text{g mol}^{-1}$. Additional error analysis for the BAT model is shown in SI Sect 5. In panels (b) and (c) of Fig. 2, we show two examples of the BAT predictions, after domain-specific optimization, compared to the AIOMFAC-generated data.

    SI Revisions: SI Sect. 5, titled "BAT Model Validation and Error Analysis", Pages 8 – 12: *copied section begins*

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.

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 $a_{w,\text{sep}}$ point. For the organic compounds in the binary aquous systems that underwent phase separation, the RMSE of BAT vs. AIOMFAC $a_{w,\text{sep}}$ predictions are listed in Table S5. Overall the BAT $a_{w,\text{sep}}$ prediction was $< \pm0.01$, the a_w prediction was $< \pm0.09$ (9 % RH), and the a_{org} prediction was $< \pm0.15$ compared to AIOMFAC.
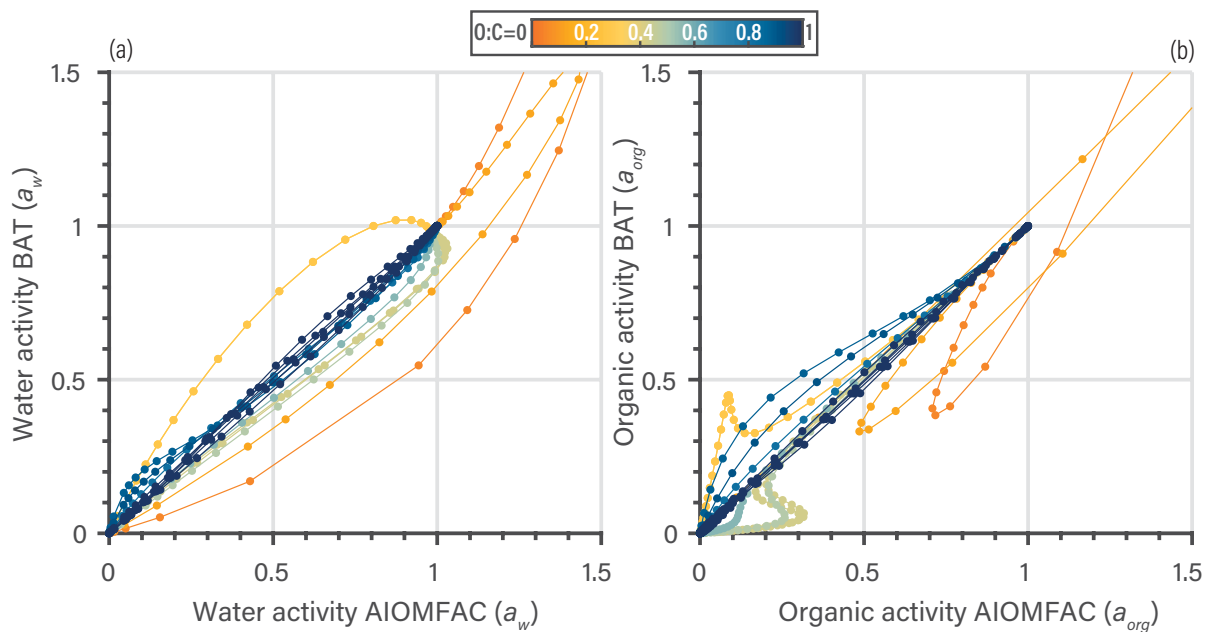
Figure 1. :**SI Figure S5** Comparisons of the BAT predictions with the AIOMFAC model predictions for the validation data set (Table S6). An activity value above one represents an unstable or metastable mixing state, and in practice the mixture would phase separate readily when given that initial mixture 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}.

Table 1. :**SI 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	Hydro-peroxide	Hydro-peroxide 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 $a_{w,\text{sep}}$	0.0066	0.0127	0.0031	0.0039	0.0061	none	0.0075	0.0032	0.0024

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; P

5 . The average error in the 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.
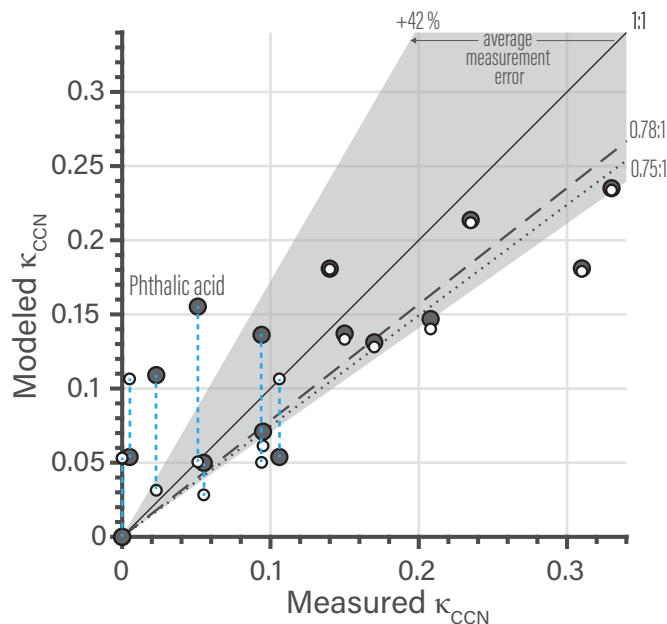
Figure 2. :SI Figure S6 Comparison of single-component organics κ_{CCN} measurements against those predicted by BAT (black circles) and AIOMFAC (white circles) model simulations of CCN activation. The blue, dashed lines connect the BAT and AIOMFAC predictions for the same species when there is a large 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, $\kappa_{CCN,BAT} = \kappa_{CCN,measured} \times 0.78 \ [\pm 0.078]$ with a Pearson's $R^2 = 0.48$. The black dotted line is the AIOMFAC model linear fit with a zero intercept, $\kappa_{CCN,AIOMFAC} = \kappa_{CCN,measured} \times 0.75 \ [\pm 0.066]$ 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

copied section ends

Note, the table and citation errors of text overflow only show up in the tracked changes text, and not in the clean/revised manuscript and SI text.

Related to this point: "Therefore, in an attempt to design a more general organic activity coefficient model, we made two important changes. First, we change the independent composition variable used in Eq. (15). Instead of mole fraction xorg, we introduce a scaled volume fraction ('org) in the series expansion of GE/RT. Second, we introduce a parameterization of the scalar c0 n coefficients by means of multivariate functions, which are dependent on common characteristics of organic molecules."

(1) Please provide some rationale why switching to volume fraction was required. It is not clear to the referee or from the text.

Authors Response: As described near the bottom of page 9: "The scaled volume fraction acknowledges that neither mole fraction nor volume fraction (nor mass fraction) perfectly accounts for the composition-dependence of activity coefficients when describing various binary systems." The switching to volume fraction is not required *per se*, but some form of scaling of the mole fraction composition scale is advantageous when the same binary Gibbs excess function is used for more than a single system and in particular when targeting a wide range in molecular masses of the organic component and therefore large shape and size differences compared to water molecules. Using volume fraction scaling is a better natural choice as it includes accounting for differences in molecular sizes, while the scaling further helps to achieve optimal model performance, which was also confirmed by preliminary tests we run before settling for the present functional form of the BAT model. We add the following text to explain this step.

Manuscript Revisions on Page 9–10: *The introduced change of composition scale improves the flexibility of this model when optimized for a wide range of binary systems characterized by the same set of model parameters ($s_1, s_2; a_{n,1}, a_{n,2}, a_{n,3}$, etc., with $n = 1, 2\ldots$). The* mole fraction scale works well for binary systems involving two components of similar molecular size and shape. However, this is rarely the case in aqueous organic mixtures with organic compounds of substantially higher molar mass than water. The volume fraction scale implicitly accounts to some extent for the size difference between organic and water molecules, which means that the coefficient functions c_n do not need to correct for the molecular size- and composition-dependence as much as when mole fraction were used. It is for a similar reason that local composition models like UNIFAC describe organic molecules as a combination of similar-sized segments (subgroups) occupying a regular lattice, which contributes to the so-called combinatorial activity in those models. The *scaled volume fraction acknowledges that neither mole fraction nor volume fraction (nor mass fraction) perfectly accounts for the composition-dependence of activity coefficients when describing various binary systems.* Alternatively, a scaled mole fraction composition scale could have been used, but we chose to scale volume fractions as the scaling coefficient values constitute a smaller adjustment when used with this composition scale, meaning that a simpler scaling function was sufficient. *Importantly, Eq. (19) remains consistent with all thermodynamic relations, including that G^E becomes zero at both limits: $\phi_{org} = 0$ (when $x_{org} = 0$), $\phi_{org} = 1$ (when $x_{org} = 1$).*

(2) I don't understand why using Redlich-Kister was selected. The RK expansion can fit to arbitrary precision. While it is true that the model is thermodynamically consistent in the limit of x = 1 and x = 0, the polynomial can lead to maxima and minima in the excess [Gibbs] function that may or may not be realistic. This is particularly concerning since the polynomial coefficients are themselves computed through a multivariate model. (It is impressive that the system converged). I am concerned that the BAT model coefficients are overfitted and not be representative of other compositions, especially with respect to a phase separation which represents a discontinuity and is highly sensitive to very small fluctuations in the excess [Gibbs] function. This issue re-emphasizes the need to independently verify the fidelity of the BAT model. When addressing this concern, the authors should discuss why RK was selected instead of the Margules or van Laar model, which would be less sensitive to error from the use of polynomials by limiting the shape of the excess [Gibbs] function. There are a few comparisons to actual data. While it is clear that the model cannot be compared to every data point in the literature, the real-world validation appears not to be systematic. For example, it's not clear why single component data from Marsh et al. (2019) was selected for hygroscopic growth and various a-pinene SOA for CCN. The composition dependence of subsaturated water content on O:C for SOA (e.g. Pajunoja et al., 2015, doi:10.1002/2015GL063142) is far more revealing than the supersaturated data. Many data sets for single subsaturated water uptake of single component organic aerosol are available. Ideally a proper validation set

Authors Response:

5   **BAT function**: We chose the Redlich–Kister functional form specifically because it could be fitted to arbitrary precision and account for extrema in activity coefficients, if necessary; however, more than two coefficient terms (c_1, c_2) seemed to add little value to the fits (see description in Section 3.2). Use of two coefficient terms (in BAT as parameterized functions) also means that the shape of the Gibbs excess function is constrained towards realistic behavior, similar to a two-parameter Margules model. We also wanted the excess Gibbs function and activity coefficients to be capable of expressing maxima and

10   minima as that behavior is important for models that allow for liquid–liquid phase separation, thus we did not use the van Laar model. Since we went with two polynomial terms and had 1000+ data points covering the range from very low to very high concentrations of aqueous organic systems to fit the model, the behaviour of BAT is well constrained in the O : C and molar mass space considered. Additional isopleth of Figure 3 for lower and higher molar masses were added to explore functional irregularities in the BAT Model Validation and Error Analysis section of the SI.

15   **Manuscript Revisions on Page 8:** *In addition G^E must be capable of expressing maxima and minima within the mixed composition space ($0 < x_{org} < 1$) to correctly capture possible phase separation behavior.*

Validations: The concerns about validation/overfitting to AIOMFAC have been addressed in the response to the first comment; validation to measurements are discussed below.

We used the comparison to measurements by Marsh et al. (2019) mainly since the chemical species in those experiments

20   are known. This is also why we did not use any measured OA data sets for subsaturated conditions, for which chemical composition is not well known. More comparisons and analyses of ambient and laboratory data sets (e.g. Pajunoja et al., 2015, doi:10.1002/2015GL063142) have been split off to future work as additional analysis is required (estimations/assumptions of volatility, molecular weight, and O : C distributions), which would distract from the the main point of introducing the VBS + BAT model. We have added a comparison plot showing modelled vs. measured κ_{CCN} data as suggested for single-

25   component (aqueous) organic aerosol.

Manuscript Revisions on Page 30–32: *After mainly comparing to data for subsaturated conditions in Fig. 11, we now focus on predictions for the regime supersaturated with respect to water vapor. In Fig. 12, the measurement derived κ_{CCN} is compared with the corresponding BAT model prediction. The data set contains 30 supersaturated droplet activation measurements of known chemical species (e.g., oleic acid, glucose, and levoglucosan). The average error in the measurements*

30   *is shown as the gray shaded area in Fig. 12, which covers the average of the κ_{CCN} range observed for each component. A subset of 18 chemicals reported a κ_{CCN} range, from which the average error was calculated to be $\pm 42\%$. The data set we used was compiled by Petters et al. (2016) and Petters and Kreidenweis (2007), which includes measurements derived from multiple sources (Broekhuizen et al., 2004; Brooks et al., 2004; Frosch et al., 2010; Huff Hartz et al., 2006; Petters et al., 2006; Petters and Kreide . Our comparison excludes the nitrogen-containing compounds. The BAT predictions assumed no organic co-condensation and*

35   *had an evolving surface tension as described in Sect. 5.3. The BAT predictions vs. measurements had an RMSE of 0.055 and overall agreed within the reported measurement error. Substantial differences are found for the $0.35 < O : C < 0.55$ range, in which the resulting κ_{CCN} is highly sensitive to a correct prediction of miscibility. For example, the miscibility is over-predicted*

for phthalic acid (O : C = 0.5) while it is under-predicted for pinic acid (O : C = 0.44), shown in Fig. 12. In the full data set of 30 molecules, another subset of 16 molecules were not in the training database of the BAT model, so a corresponding plot with only this validation data is shown in the section 5.1 of the SI, including predictions by both BAT and AIOMFAC. The validation data shows similar agreement to Fig. 12, with a measurement vs. BAT RMSE of 0.061 and measurement vs. AIOMFAC RMSE

5   *of 0.059. The AIOMFAC κ_{CCN} predictions are better in the miscibility transition region than the BAT model, but overall the models show similar predictive skill for this metric. We chose to focus on well-defined chemical systems for all of the direct BAT model–measurement comparisons, allowing for minimal uncertainty in the input data. Additional comparisons of BAT to complex ambient and laboratory OA systems will be carried out in the future, since additional analyses are necessary for the estimation of volatility, molecular mass, and O : C distributions. Such analyses will enable a fair evaluation of VBS + BAT*

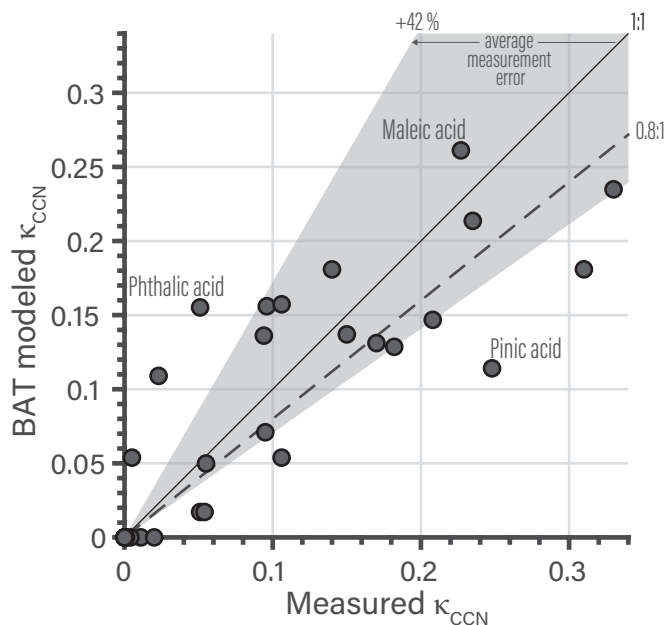10   *model predictions against measurements for systems that are unresolved on the molecular composition level.*

Figure 3. :**Main text Figure 12:** Single-component organic aerosol measurements of κ_{CCN} are compared against those predicted by corresponding BAT model simulations of CCN activation. The gray shading represents $\pm 42\,\%$ average uncertainty in the measured κ_{CCN}. The dashed line is a linear fit with a zero intercept, $\kappa_{CCN,BAT} = \kappa_{CCN,measured} \times 0.799\,[\pm 0.059]$ with a Pearson's R^2 of 0.66. The model–measurement RMSE was 0.055. The BAT simulations assume a 100 nm diameter equivalent volume of organic matter at the CCN activation point. The droplet surface tension is calculated as a volume-weighted mean. A list of the 30 measurement points is given in Table S6 of the SI, with the data obtained from the following studies: Broekhuizen et al. (2004); Brooks et al. (2004); Petters et al. (2006); Petters and Kreidenweis (2007); Petters et al. (2009, 2016); Frosch et al. (2010); Hu

Other comments

The tone of the hygroscopic growth and CCN section should be revised. For examples

"Over the past decade, the research community then progressed by characterizing (organic) aerosol hygroscopic growth measurements by a single κ value, with sometimes inconsistent distinction between a κ value at subsaturated and supersaturated humidity conditions."

"Our clear distinction between κ CCN and the more general κHGF helps the community understand clearly the subsaturated and supersaturated behavior of organic aerosol"

While it is true that there has been a debate on κCCN and κHGF the authors should acknowledge that 100s of experimental and modeling papers were devoted to this subject, with many important individual contributions explaining the origin of the discrepancy and the composition dependence of κHGF. While the BAT model may capture some of these now very well understood effects, it does not really reveal anything new. Please rephrase the text and/or provide a more nuanced perspective on the topic.

Authors Response: Right, the BAT model does not reveal any new processes. Nevertheless, it may provide a different way to visualize the sub- vs. supersaturated hygroscopicity signatures. Everything we showed could also be done – and most has been done – with models like UNIFAC or AIOMFAC. We have added that point and revised this section's statements to address this reviewer's concerns. An advantage of the BAT model, with its intrinsic and continuous dependence on M_{org} and O : C, is its ability to compute the isolines shown in Figure 10, which would have to be discretized by a set of molecular formulas in a similar figure when using UNIFAC/AIOMFAC.

Manuscript Revisions on Page 25–27:

copied section begins

Our last model application focuses on κ at the CCN activation point, denoted as κ_{CCN} of the organic aerosol. The BAT model is used to understand composition effects on the hygroscopic growth parameter of organic species at CCN activation conditions and the related ongoing discussion within the atmospheric science community. The BAT model can predict an entire Köhler curve directly and does not rely on a κ_{CCN} prediction for applications in the context of cloud droplet formation thermodynamics. Thus, the exercise of predicting κ_{CCN} is here mainly carried out to inform on the relationship with existing approaches. The κ-Köhler framework reduces hygroscopic growth to a single parameter (κ) that can be used to compare the properties of different potential CCN particles (Petters and Kreidenweis, 2007). Over the past decade, the research community then progressed by characterizing (organic) aerosol hygroscopic growth measurements by a single κ value , with sometimes inconsistent distinction between a κ value at subsaturated and supersaturated humidity conditions for ease of comparison and use for parameterizations of CCN activation in large-scale models (e.g. Petters and Kreidenweis, 2007; Rastak et al., 2017). The overarching goal was to link measured aerosol physicochemical properties to CCN activation behavior (critical supersaturation, critical dry diameter, etc.). A common approach was to fit a linear dependence of κ to organic O : C (Jimenez et al., 2009; Chang et al., 2010; Massoli et al., 2010; Cappa et al., 2011; Duplissy et al., 2011; Frosch et al., 2011; Lambe et al., 2011; Wong et al., 2011; Rickards et al., 2013; Thalman et al., 2017). A resulting linear fit was not always consistent with observations, due to the nonlinear behavior of κ vs. O : C, so Kuwata et al. (2013) introduced a set of water-solubility bins to account for nonlinear step changes. More recently, Wang et al. (2019) focused on relating κ to molar mass and assumed ideal mixing of organics with water. There are at least two main problems that likely led to the current understanding of κ for organic aerosol factors

that many of the previous approaches do not fully account for. The first ~~was the common assumption of~~ is assigning a single κ value and assuming it to be representative at all RH levels, which ~~is often inaccurate (~~has been shown to be inaccurate in multiple cases as this treatment does not account for non-ideal behavior changing with RH (or a_w), especially in the RH range of 90 – 100 % (see Fig. 8b). The second being the ~~fixation with using~~ use of a linear function to describe ~~nonlinear behavior .~~

5 ~~We may gain a better insight on the link between organic aerosol properties to CCN activity properties by using~~ the non-linear behavior caused by liquid–liquid phase separation. More advanced thermodynamic models, like UNIFAC and AIOMFAC, have been used to gain insight into the complex CCN activation process accounting for phase separation and non-ideal mixing (Petters et al., 2016; Ovadnevaite et al., 2017; Renbaum-Wolff et al., 2016; Rastak et al., 2017; Hodas et al., 2016). The BAT model can simulate the same processes as those more detailed thermodynamic models, but with less (or incomplete) information about

10 the molecular structure and/or composition of the organic aerosol fraction. We acknowledge that there remain a number of challenges accompanying predictions of CCN activation potential, including accounting for composition-dependent bulk–surface partitioning of different organic and inorganic components in multicomponent aerosol and associated evolving surface tension (e.g. Ruehl et al., 2016; Malila and Prisle, 2018; Davies et al., 2019). At present, those aspects may be best understood and represented by detailed process models, though future BAT extensions may enable improvements also on a reduced-complexity

15 level.

    The reduced-complexity inputs of the BAT model ~~to predict the~~ and its continuous behavior as a function of $O:C$ and M_{org} ~~dependence of κ_{CCN}. In this context~~ allow for establishing a direct link between those organic aerosol properties ($O:C$ and M_{org}) and the predicted CCN activation potential. For these BAT model predictions, we revert to the original definition of κ_{CCN} by assuming no organic co-condensation in Eq. 22 (i.e., $V_{org,\mathrm{dry}} = V_{org}$). Accounting for the Kelvin effect with an

20 assumption about the air–droplet surface tension, one can calculate the equilibrium saturation ratio S of the aerosol / CCN,
    copied section ends

Manuscript Revisions on Page 34: *The ~~the~~ interplay between $O:C$, molar mass, and water uptake for CCN activation ~~revealed~~ clearly show the complex behavior of organic κ_{CCN} values. Our ~~clear~~ distinction between κ_{CCN} and the more general κ_{HGF} helps ~~the community understand clearly the~~ to differentiate between the subsaturated and supersaturated behavior of*

25 *organic aerosol. The use of the BAT model in κ_{CCN} prediction correctly captures the nonlinear dependence of κ_{HGF} (and κ_{CCN}) on organic properties and is preferable to previous linear fits~~using~~ .*

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.

Cappa, C. D., Che, D. L., Kessler, S. H., Kroll, J. H., and Wilson, K. R.: Variations in organic aerosol optical and hygroscopic properties upon heterogeneous OH oxidation, J. Geophys. Res., 116, 1–12, https://doi.org/10.1029/2011JD015918, 2011.

Chang, R. Y., Slowik, J. G., Shantz, N. C., Vlasenko, A., Liggio, J., Sjostedt, S. J., Leaitch, W. R., and Abbatt, J. P.: The hygroscopicity parameter (κ) of ambient organic aerosol at a field site subject to biogenic and anthropogenic influences: Relationship to degree of aerosol oxidation, Atmos. Chem. Phys., 10, 5047–5064, https://doi.org/10.5194/acp-10-5047-2010, 2010.

Davies, J. F., Zuend, A., and Wilson, K. R.: Technical note: The role of evolving surface tension in the formation of cloud droplets, Atmos. Chem. Phys., 19, 2933–2946, https://doi.org/10.5194/acp-19-2933-2019, 2019.

Duplissy, J., DeCarlo, P. F., Dommen, J., Alfarra, M. R., Metzger, A., Barmpadimos, I., Prevot, A. S. H., Weingartner, E., Tritscher, T., Gysel, M., Aiken, A. C., Jimenez, J. L., Canagaratna, M. R., Worsnop, D. R., Collins, D. R., Tomlinson, J., and Baltensperger, U.: Relating hygroscopicity and composition of organic aerosol particulate matter, Atmos. Chem. Phys., 11, 1155–1165, https://doi.org/10.5194/acp-11-1155-2011, 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.

Frosch, M., Bilde, M., DeCarlo, P. F., Jurányi, Z., Tritscher, T., Dommen, J., Donahue, N. M., Gysel, M., Weingartner, E., and Baltensperger, U.: Relating cloud condensation nuclei activity and oxidation level of α-pinene secondary organic aerosols, J. Geophys. Res. Atmos., 116, 1–9, https://doi.org/10.1029/2011JD016401, 2011.

Hodas, N., Zuend, A., Schilling, K., Berkemeier, T., Shiraiwa, M., Flagan, R. C., and Seinfeld, J. H.: Discontinuities in hygroscopic growth below and above water saturation for laboratory surrogates of oligomers in organic atmospheric aerosols, Atmos. Chem. Phys., 16, 12 767–12 792, https://doi.org/10.5194/acp-16-12767-2016, 2016.

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.

Jimenez, J. L., Canagaratna, M. R., Donahue, N. M., Prevot, a. S. H., Zhang, Q., Kroll, J. H., DeCarlo, P. F., Allan, J. D., Coe, H., Ng, N. L., Aiken, a. C., Docherty, K. S., Ulbrich, I. M., Grieshop, a. P., Robinson, a. L., Duplissy, J., Smith, J. D., Wilson, K. R., Lanz, V. a., Hueglin, C., Sun, Y. L., Tian, J., Laaksonen, A., Raatikainen, T., Rautiainen, J., Vaattovaara, P., Ehn, M., Kulmala, M., Tomlinson, J. M., Collins, D. R., Cubison, M. J., Dunlea, E. J., Huffman, J. a., Onasch, T. B., Alfarra, M. R., Williams, P. I., Bower, K., Kondo, Y., Schneider, J., Drewnick, F., Borrmann, S., Weimer, S., Demerjian, K., Salcedo, D., Cottrell, L., Griffin, R., Takami, A., Miyoshi, T., Hatakeyama, S., Shimono, A., Sun, J. Y., Zhang, Y. M., Dzepina, K., Kimmel, J. R., Sueper, D., Jayne, J. T., Herndon, S. C., Trimborn, a. M., Williams, L. R., Wood, E. C., Middlebrook, a. M., Kolb, C. E., Baltensperger, U., and Worsnop, D. R.: Evolution of organic aerosols in the atmosphere., Science, 326, 1525–1529, https://doi.org/10.1126/science.1180353, 2009.

Kuwata, M., Shao, W., Lebouteiller, R., and Martin, S. T.: Classifying organic materials by oxygen-to-carbon elemental ratio to predict the activation regime of Cloud Condensation Nuclei (CCN), Atmos. Chem. Phys., 13, 5309–5324, https://doi.org/10.5194/acp-13-5309-2013, 2013.

Lambe, A. T., Onasch, T. B., Massoli, P., Croasdale, D. R., Wright, J. P., Ahern, A. T., Williams, L. R., Worsnop, D. R., Brune, W. H., and Davidovits, P.: Laboratory studies of the chemical composition and cloud condensation nuclei (CCN) activity of secondary organic aerosol

(SOA) and oxidized primary organic aerosol (OPOA), Atmos. Chem. Phys., 11, 8913–8928, https://doi.org/10.5194/acp-11-8913-2011, 2011.

Malila, J. and Prisle, N. L.: A monolayer partitioning scheme for droplets of surfactant solutions, J. Adv. Model. Earth Syst., 10, 3233–3251, https://doi.org/10.1029/2018MS001456, 2018.

5 Massoli, P., Lambe, A. T., Ahern, A. T., Williams, L. R., Ehn, M., Mikkilä, J., Canagaratna, M. R., Brune, W. H., Onasch, T. B., Jayne, J. T., Petäjä, T., Kulmala, M., Laaksonen, A., Kolb, C. E., Davidovits, P., and Worsnop, D. R.: Relationship between aerosol oxidation level and hygroscopic properties of laboratory generated secondary organic aerosol (SOA) particles, Geophys. Res. Lett., 37, 1–5, https://doi.org/10.1029/2010GL045258, 2010.

Ovadnevaite, J., Zuend, A., Laaksonen, A., Sanchez, K. J., Roberts, G., Ceburnis, D., Decesari, S., Rinaldi, M., Hodas, N., Facchini, M. C.,
10 Seinfeld, J. H., and O' Dowd, C.: Surface tension prevails over solute effect in organic-influenced cloud droplet activation, Nature, 546, 637–641, https://doi.org/10.1038/nature22806, 2017.

Petters, M. D. and Kreidenweis, S. M.: A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, Atmos. Chem. Phys., 7, 1961–1971, https://doi.org/10.5194/acp-7-1961-2007, 2007.

Petters, M. D., Kreidenweis, S. M., Snider, J. R., Koehler, K. A., Wang, Q., Prenni, A. J., and Demott, P. J.: Cloud droplet activation
15 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
20 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.

Rastak, N., Pajunoja, A., Acosta Navarro, J. C., Ma, J., Song, M., Partridge, D. G., Kirkevåg, A., Leong, Y., Hu, W. W., Taylor, N. F., Lambe, A., Cerully, K., Bougiatioti, A., Liu, P., Krejci, R., Petäjä, T., Percival, C., Davidovits, P., Worsnop, D. R., Ekman, A. M., Nenes,
25 A., Martin, S., Jimenez, J. L., Collins, D. R., Topping, D. O., Bertram, A. K., Zuend, A., Virtanen, A., and Riipinen, I.: Microphysical explanation of the RH-dependent water affinity of biogenic organic aerosol and its importance for climate, Geophys. Res. Lett., 44, 5167–5177, https://doi.org/10.1002/2017GL073056, 2017.

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

30 Renbaum-Wolff, L., Song, M., Marcolli, C., Zhang, Y., Liu, P. F., Grayson, J. W., Geiger, F. M., Martin, S. T., and Bertram, A. K.: Observations and implications of liquid–liquid phase separation at high relative humidities in secondary organic material produced by α-pinene ozonolysis without inorganic salts, Atmos. Chem. Phys., 16, 7969–7979, https://doi.org/10.5194/acp-16-7969-2016, 2016.

Rickards, A. M., Miles, R. E., Davies, J. F., Marshall, F. H., and Reid, J. P.: Measurements of the sensitivity of aerosol hygroscopicity and the κ parameter to the O/C ratio, J. Phys. Chem. A, 117, 14 120–14 131, https://doi.org/10.1021/jp407991n, 2013.

35 Ruehl, C. R., Davies, J. F., and Wilson, K. R.: An interfacial mechanism for cloud droplet formation on organic aerosols, Science, 351, 1447–1450, https://doi.org/10.1126/science.aad4889, 2016.

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.

Thalman, R., De Sá, S. S., Palm, B. B., Barbosa, H. M., Pöhlker, M. L., Alexander, M. L., Brito, J., Carbone, S., Castillo, P., Day, D. A., Kuang, C., Manzi, A., Ng, N. L., Sedlacek, A. J., Souza, R., Springston, S., Watson, T., Pöhlker, C., Pöschl, U., Andreae, M. O., Artaxo, P., Jimenez, J. L., Martin, S. T., and Wang, J.: CCN activity and organic hygroscopicity of aerosols downwind of an urban region in central Amazonia: Seasonal and diel variations and impact of anthropogenic emissions, Atmos. Chem. Phys., 17, 11 779–11 801, https://doi.org/10.5194/acp-17-11779-2017, 2017.

Wang, J., Shilling, J. E., Liu, J., Zelenyuk, A., Bell, D. M., Petters, M. D., Thalman, R., Mei, F., Zaveri, R. A., and Zheng, G.: Cloud droplet activation of secondary organic aerosol is mainly controlled by molecular weight, not water solubility, Atmos. Chem. Phys., 19, 941–954, https://doi.org/10.5194/acp-19-941-2019, 2019.

Wong, J. P., Lee, A. K., Slowik, J. G., Cziczo, D. J., Leaitch, W. R., MacDonald, A., and Abbatt, J. P.: Oxidation of ambient biogenic secondary organic aerosol by hydroxyl radicals: Effects on cloud condensation nuclei activity, Geophys. Res. Lett., 38, 1–6, https://doi.org/10.1029/2011GL049351, 2011.