We thank both reviewers for their constructive comments which we think have helped us improve the manuscript. Our point-by-point responses to both reviewer comments along with the corresponding changes we have made to the revised manuscript (attached in the end of the letter with the changes highlighted) are given below. Besides the revisions requested by the reviewers we have modified Sect. 2.2.4 according to the corrigendum we published as an author comment, as well as corrected some typos and small inclusions we found in the original manuscript.

Response to Reviewer #1

We thank this Reviewer for the thoughtful read of our manuscript and the insightful comments. Our point-by-point responses to the Reviewer’s comments are below. The direct quotes from the Reviewer are in italics and our responses in normal font.

(1) I am very unsure of, and uncomfortable with, the use of the terminology Basis Set in the context of the current study. It makes no sense in terms of the common usage in theoretical and computational chemistry, nor that in linear algebra. Its usage in terms of the well-established VBS was also difficult to understand and its adoption in the current work by analogy to the VBS requires explanation. As stated, the current manuscript considers "a continuous distribution of solubilities" in classes one decade apart. This provides a suitable reference scale on an appropriate axis, but I do not see where there is any reference "set" (of functions, of compounds, of properties, of vectors etc...) that can be used to make sense of the impacts of particle component solubility. This is particularly important in the context of the non-independence of the presented component description in terms of solubility and the extended volatility description where the second dimensions considers the degree of oxygenation (in terms of O:C or oxidation state), see below.

We understand the point of the reviewer and indeed the different dimensions of the “basis sets” are not expected to be independent of each other. We will therefore primarily use the term "solubility distribution" instead of "solubility basis set (SBS)" throughout the revised manuscript, except for the point in which the analogy to the widely-used VBS approach is discussed.

(2) To expand on this point, given the relationship between aqueous solubility of organic molecules and their polarity (and hence practically, their oxygen-containing functionality), the solubility consideration in the current paper is clearly not independent of the 2-D VBS of Donahue et al., 2011 or the Carbon number - oxidation state representation of Kroll et al., 2011. Whilst both approaches are mentioned in the current work, the relationship between the approaches is insufficiently discussed. It appears that the current work concerns itself with the cloud droplet activation behaviour of OA components, whereas the "VBS" approaches are more concerned with formation and transformation of the OA. However, the relationship between degree of oxygenation and hygroscopicity has been widely investigated, both in terms of the VBS and otherwise (e.g. in terms of AMS m/z 44 in the paper introducing the 2D-VBS). Introduction of yet another approach without contextualisation appears to add to the confusion, rather than to its clarification.

It is expected that the mixture solubility distribution will depend on both polarity and thus O:C ratio (Donahue et al., 2011; Kroll et al., 2011) as well as the molar mass (Shiraiwa et al., 2014) of the mixture constituents. To illustrate this, we have added a figure depicting an example solubility distribution for a representative case of α-pinene SOA (Chen et al., 2011), where the solubilities have been estimated using the SPARC (see e.g. Wania et al., 2014 and references therein) prediction tool. Also the mean molecular mass and O:C ratio for each solubility bin are presented. A brief discussion of the expected dependence of the solubility on these different molecular properties has been added along with the description of the new figure to the revised manuscript.

(3) Line 7 p28530 it is stated that "This assumption is justified to a first order by the different equilibration time scales of the droplets with respect to water vapour and the organic vapours in typical
The moist atmosphere is a ration ratio by cooling (updraught in a cloud, radiative or advective in a fog). The (low) rate of change of saturation ratio caused by this will lead to the mass flux of water. Since the number of water molecules is high, it is frequently assumed that the number of collisions leads much more rapid condensation of water vapour than of other components. However, the rate of change of saturation ratio of organic molecules can be very much higher (owing to both temperature, but equally or more importantly photochemical reaction). Since the vapour pressure of the organic can be many orders of magnitude lower than water, a strong diffusion gradient between the gas phase and the particle surface can be very rapidly established. There are very many potential oxygenated organic compounds that can all be rapidly produced in response to strong emission and photochemical changes. It is far from clear to me that the addition of soluble mass to a particle in moving towards cloud base can be ignored to a first order approximation (and hence that the solubility distribution of components at lower RH will be the same as the solubility distribution of components close to activation). The Topping et al. Nature Geoscience paper (doi:10.1038/hge01809) should be referenced in this regard. Again, the blurred relationship between volatility treated in the "VBS" approaches and solubility treated here is related to this. It appears conceptually peculiar to silo treatments of partitioning and water uptake into the separate approaches; the moist atmosphere is a single entity in which both processes simultaneously occur.

We agree that this is an important assumption and that it deserves additional discussion. It is also true that for a comprehensive picture of the CCN activation, all the relevant dynamic processes for all the involved species (and their gas phase concentrations) need to be explicitly considered. This requires, however, a much more complex modeling framework and detailed information about the connections between volatility and water solubility of the different organic compounds that is not currently available. For the scope of this work we felt that it is important to focus on the solubility effects on the CCN activation process, without adding the uncertainty related to e.g. the gas-phase concentrations of the organic species and atmospheric dynamics. This is clearly a first step and investigations with a cloud parcel model including all the relevant dynamic processes will be carried out in a future study. We think that this approach is justified, as this is the first study investigating the sensitivity of the CCN activation processes to a wide selection of solubility distributions. In the revised manuscript we have added a reference to Topping et al. (2012) and replaced the statement of the justification of this assumption with a more nuanced discussion of the corresponding issues mentioned by the reviewer.

(4) I’m a little confused in how the approach described here to calculate the equilibration of components between the insoluble core and the surrounding aqueous shell relates to previous approaches to deal with phase separation in complex mixtures in atmospheric aerosol (notably that presented by Topping et al., 2013, doi:10.1039/C3FD00047H, which should be referenced). The solubility in the non-aqueous and aqueous phases must implicitly account for non-ideality (and e.g. degree of dissociation of weak acids and bases) and hence must be considered an effective solubility in the mixture. In the aqueous phase, this effective solubility may or may not be relatively simply related to the infinite dilution solubility of the component. In any case, it will very likely be relative humidity dependent. As such, the effective solubility distribution in the two phases will be RH dependent and this will (potentially significantly) affect the shape of the Kohler curves and the results illustrated in Figure 3. Coupled to the equilibration between the gas phase and the two condensed phases, it might be expected that this effect is substantial. In figure 8 a simple sensitivity to an assumed activity relationship is explored (using equation 5). Such calculations can be carried out for simple real mixtures where the activity coefficient data (albeit on a mole fraction scale) are available (and complex mixtures where the activity coefficients can be predicted). This sensitivity is assumed to constrain the real behaviour (very likely correct). It would be useful to know whether the real behaviour more normally follows one or the other of these limiting cases. Even in the absence of confirmatory experimental data, such calculations are possible.
We have added the reference to Topping et al. (2013) and the corresponding discussion. We have also cited the work of Cappa et al. (2008) with some discussion on the potential non-ideality of multicomponent mixtures of dicarboxylic acids. Unfortunately, there is very little experimental information available about atmospheric organic compound activity coefficients in the aqueous or multicomponent organic phases. Data is available for individual organic compounds, but it is not clear how useful these data would be to estimate the mixture behavior. Our simplified ideality assumption for the aqueous phase activities below the saturation concentrations is a reasonable starting point. The corresponding solubilities discussed here should be interpreted, as the reviewer suggests, as “effective” solubilities. A brief discussion has been added to address this point in the revised manuscript.

(5) It is stated on line 15 of 28532 that "... all the equations presented below can be rederived in a relatively straightforward manner taking into account a potential difference between the mole and the mass fractions in the organic phase". This is correct, but the reference scale difference for the activity coefficients (mass or mole fraction) must be emphasised. This must be made explicit - it is the same problem as encountered in the VBS and is more important here, since the molecular weight distribution and degree of functionalisation will significantly affect effective solubility and hence phase partitioning.

This is correct. We have added a sentence to the revised manuscript emphasizing the fact that we have studied the special case where the mole and mass fraction scales are the same, but that these should be distinguished for realistic atmospheric mixtures.

Minor:

(6) Some reference to the expected individual component effective solubility range for atmospheric components would be helpful. The very heavy functionalisation (of multifunctional peroxides, peroxycids etc...) expected and required to attain observed O:C ratios with a measured molecular weight distribution of organics provides some surprisingly high solubilities. Along with the very low solubilities of slightly functionalised high molecular weight primary compounds, this will lead to a very wide distribution of solubilities as shown in Figure 2. Are there any data available to say whether the real atmosphere is likely to be closer to one or another of the chosen distributions? If not, what are the difficulties and are there suggested routes forward?

To link our approach better to realistic atmospheric organic mixtures, we have now constructed the solubility distribution for an example composition of secondary organic aerosol (SOA) mixture derived from α-pinene oxidation (see Chen et al., 2011 for details of the gas-phase composition and conditions corresponding to this distribution). It should be noted though that this is only one example, and the organic aerosol formed under different conditions or derived from other precursors can have vastly different solubilities, nevertheless being equally well represented by the solubility distribution approach. The results from our manuscript can then further be used to assess the requirements (e.g., range of solubilities, resolution) of a solubility distribution measurement approach to represent the CCN activation of these mixtures. We have added discussion on the link of the study to realistic atmospheric organic mixtures, along with adding a new figure depicting an estimated realistic solubility distribution to the revised manuscript.

(7) I am not sure that the 2 sentences in the abstract from line 17 to 21 and 21 to 24 are not contradictory. Please check. Can the stated material in the first sentence be assumed completely soluble and is this material not just a subset of the material considered within the distribution mentioned in the second sentence and hence requiring understanding of the details of its distribution in solubility? In any case, the arguments considered together in such close proximity before the reader has accessed the rest of the paper are too complicated for easy interpretation.

Indeed these two statements can cause some confusion. We have revised the statements to read:
"Depending on what was assumed about the organic phase activity, material with solubilities larger than 0.1-100 g L\(^{-1}\) could be treated as completely soluble in the CCN activation process over atmospherically relevant particle diameters and supersaturations. Our results indicate that understanding the details of the solubility distribution is thus probably necessary only in the range of 0.1 to 100 g L\(^{-1}\), while resolution outside this solubility range will not add much knowledge to understanding the CCN activation of the mixture."

(8) On p28526, 3 prior assumptions are described that are considered in the paper. The kappa consideration is not the only single parameter representation of hygroscopicity (e.g. Wex et al., 2007, doi:10.1029/2006GL028260), nor the first (e.g. Rissler et al., ACP, 2004). These treatments should be acknowledged.

Thank you for pointing this out, we will correct this in the revised manuscript.

(9) It isn’t clear how the presence of inorganics effects results in this paper. It is likely that there will be an inorganic fraction under most expected atmospheric conditions and this should be mentioned in the discussion.

This is an important point as the present work addresses explicitly only organic particles. The work can be extended to include inorganic salt adding one more variable for consideration (the organic mass fraction in the particle). However, accurate considerations on the mixture activity and potential phase-separation effects will become much more complex, requiring detailed experimental data as well as advanced thermodynamic models – being thus a very interesting topic for a future study. We will add a brief discussion on this to the revised version of the manuscript.

References


**Response to Reviewer #2**

We thank the reviewer for the constructive comments. Our point-by-point responses to the Reviewer’s comments are below. The comments of the Reviewer are in *italics* and our responses in normal font.

1) The idea of the solubility basis set is novel and promising. It excites interest in the possibility to develop accurate computationally efficient parameterizations of the evolution of solubility for large scale models. The study at hand presents only limited calculations that probe parts of the parameter space.

The proposed treatment of the organic aerosol solubility distribution does simplify the multiphase thermodynamics. However, we have covered most of the atmospherically relevant parameter space related to the coupling of organic mixture solubility and CCN activation, namely:

1) the characteristics of atmospherically relevant solubility distributions (number of compounds, solubility range, shape of distribution and assumption about organic phase activity);
2) the sizes of the particles that can act as CCN;
3) atmospheric supersaturations.

While it is true that we did not cover the range of other thermodynamic properties (e.g., densities, molar masses as well surface tensions) of the organics, it was a conscious choice to isolate the solubility-related effects, which we chose to be the scope of this study – in addition to the introduction of the solubility distribution framework. Including the variation of all the other relevant thermodynamic properties of the organic compounds would be an interesting topic for future study. Furthermore, as also pointed out by Reviewer #1, to fully assess the impact of solubility on cloud formation, the results presented here need to be coupled to a model accounting for gas-particle and surface-bulk partitioning of the organics as well as atmospheric dynamics and the condensation kinetics of water vapor – framing out another direction for future work. We have clarified the justification for the chosen scope and added discussion about the potential future directions along these lines in the revised manuscript.

2) Although the proposed treatment seems to capture the important physics, there are errors in the theory that require correction.

The major issue is related to the use of the solution density in the Kelvin term (see below). We have now repeated the calculations according to the suggestions of the reviewer and found that its impact on the presented results is small and does not affect the conclusions of the study. The manuscript has been revised accordingly.

3) The presented main finding that is advertised - knowledge of solubility in the range of 0.1 and 100 g L⁻¹ is most critical for modeling CCN - is already well known from past studies (e.g. Hori et al., 2003, *J. Aerosol Sci.*) and insufficient to merit publication of a new paper on the subject.

The Hori et al. (2003) study and a number of similar previous studies have focused on a small number of specific compounds – thus investigating only specific points in the parameter space. To our knowledge, previous studies have not systematically explored such a comprehensive set of possible solubility distributions and atmospheric conditions, including the consequences of mixture effects on solubility. This is of course something that is only achievable through theoretically generated mixtures like ours, but is also among the main strengths of our approach. Additional important contributions of our work include:
1) The introduction of the solubility distribution as a framework for representing complex organic aerosol solubility.

2) The finding that material with effective solubility above 0.1-100 g L\(^{-1}\) behaved as completely soluble for most of the approximately 6000 unique combinations of solubility distributions (representing 144 unique mixtures), particle sizes and supersaturations studied. The most important piece of information to know in this regard is the thermodynamics of the insoluble organic phase.

3) For the large range of organic mixture solubilities, particle dry diameters and supersaturations explored, the single-parameter approaches (using single \(\kappa\) or \(\varepsilon\) value to describe the solubility of the mixture) were generally sufficient to predict the activation diameters of the mixture aerosols within 10%.

4) Our results also contribute to unraveling the reasons behind the "apparent simplicity" of atmospherically-relevant complex organic mixtures (e.g. Engelhart et al., 2011). Based on our results it is clear that even with vastly different solubility distributions one can yield very similar CCN-activation behavior (and consequently values of \(\kappa\) or \(\varepsilon\)), as the parameter that matters is the material above \(c_t\). This result is perhaps not surprising but it has not been systematically probed for such a large number of unique mixtures.

We have modified the abstract and conclusions of the revised manuscript to better highlight these findings and their novelty in relation to previous work, adding also reference to the work by Hori et al., 2003.

4) The evaluation of the full model against the ideal mixture, kappa-model and epsilon-model seem not very relevant to me, partially attacking and deconstructing a strawman.

We respectfully disagree, although we understand that we are essentially comparing what is supposed to be a simplification of the more complex theory to the original (more complex) description. First of all, we quantitatively assess how large error in the activation diameter one makes in assuming these simplified solubility descriptions instead of the "full" picture, putting the importance of solubility into the context of the other uncertainties related to cloud-formation process. The results convincingly show that in many cases, only one parameter (either \(\kappa\) or \(\varepsilon\)) is definitely enough to accurately calculate the activation diameter in e.g. climate models. The error introduced by this simplification is negligible compared to other uncertainties. Second, while the results make physical sense, we do not think that they can be deduced without a systematic study like ours. For instance, the importance of the assumption on the organic phase interactions – which turns out to be quite important in defining \(c_t\) (and thus highly relevant for linking e.g. experimentally determined \(\varepsilon\) or \(\kappa\) to real mixture properties) at given conditions – is something that has not been clearly pointed out before.

5) Instead, the most pressing question, how to effectively model the 50% point of partitioning (\(c_t\)) in relationship to the aqueous phase formed by a mix of species with different hygroscopicitics (including inorganic compounds and the dependence of the number of components) is not at all or insufficiently explored. Not varying the molecular weight and not including inorganic compounds seems to miss the most important question: how does the aqueous phase (number of substances and their hygroscopicity) influence \(c_t\) for a given solubility basis set? If inorganic or higher hygroscopic organics are present in the mix, \(c_t\) should shift to lower values. In my opinion, quantifying shifts in \(c_t\) in the context of the SBS and hygroscopicity of the aqueous phase ought to be the main focus of this work.

Indeed, as also indicated by the distributions present in Fig. 9 in the present paper, the \(c_t\) depends also on the number of components, supersaturation, particle size and importantly what is assumed about the organic phase thermodynamics. We have added a brief discussion of these points to the revised manuscript as well as a new figure illustrating the dependence of \(c_t\) on the number of components present in the mixture as well as the molar masses of these species. These new results show that what by far matters the most is the assumption about the organic phase activity. In the case of the unity activity
assumption, the dependence on the number of components is in practice the same behavior as observed in the distributions presented in Fig. 8 of the original manuscript.

While it would be interesting to do a specific study on the impact of inorganic compounds on \( c_t \), we feel that it is out of the scope of this paper, whose target is to lay out the general framework and behavior of a large number of different solubility distributions rather than look at any specific mixtures. To the extent that we think it fits the scope of this work, we believe that the solubility distributions that have been probed also represent well the case where an inorganic component would be present – particularly as combined with demonstrating the sensitivity of the results to the molar mass. Furthermore, we have added discussion on the dependence on the \( c_t \) on molar mass and O:C ratio through including an example of atmospherically relevant organic aerosol mixture accompanied by a new figure.

(6) I suggest to include a discussion paragraph on the vision how the solubility basis may be useful beyond the immediate sensitivities that are being explored in this manuscript. It may be useful to explore this in the context of experimental data showing that the OA hygroscopicity increases are driven by evolving solubility with chemical aging as discussed in Kuwata et al., 2012, ACP and Suda et al., 2014 ES&T. This discussion should include computational considerations. For example, what is the current (or expected gain) from running a 10, 100, 1000 component solubility model vs a model that has could predict ct, and kappa for the aqueous phase. I performed a quick test running a solubility model with unique solubility, molecular weight, and mixing fraction and CPU time only increased when reaching 10000 unique components. Presumably the purpose the SBS to be (much) more computationally efficient and should comment and factors influencing computational speed.

This is a good idea and we have added the corresponding brief discussion about the future use of the proposed framework to the revised manuscript. We also believe that the solubility distribution can be a useful approach for modeling how the mixture hygroscopicity evolves with atmospheric age, linking it to the chemical properties of the molecules present in the mixture. In this regard, we have added references to the work by Kuwata et al. (2012) and Suda et al. (2014). Furthermore, as also stated above, we believe that the solubility distribution framework can yield useful insights in the interpretation of laboratory data for complex organic mixtures – in particular on the apparent simplicity of their CCN activation despite the well-known molecular complexity of the mixtures. Furthermore, the solubility distribution approach will probably be useful in studies investigating the relative importance of solubility versus semi-volatile partitioning (e.g. Topping et al., 2012) and adsorption effects (e.g. Kumar et al., 2009) for water uptake and CCN activation.

We have also added a brief discussion on the computational cost to the revised manuscript. Indeed, 10000 compounds is the limit where significant increase in the required computational time is expected for the simple Köhler-theory -based CCN-activation calculation. However, for large-scale models, every additional simulated species (in this case solubility bin) can increase the computational cost significantly due to the need of simulating its atmospheric chemistry, transport, removal, etc. The resolution needed for accurate treatment will require better understanding of the actual solubility distributions and their linkages with other parameters (e.g., volatility, O:C, etc.).

(7) Although there is nothing really wrong with it per se, the adoption of mass units deviates from what has been done in the past and seems awkward to me. It doesn’t make the theory novel and may lead to unnecessary confusion when comparing against other manuscripts. The new part is the solubility basis set on which the focus should lie.

We assume that the Reviewer is referring to the use of mass units in expressing the solubilities – as for the mixture composition we have chosen a case where the mass and the molar units are in fact equivalent. To our knowledge expressing pure-component solubilities in g L\(^{-1}\) (or in mass per mass of water) is standard in e.g. chemical engineering handbooks, so it is difficult to know what (and which past work) the Reviewer is referring to. We would therefore prefer to keep the current units and also add some text clearly acknowledging the difference between molar and mass-based units.
Eq. (1): The use of solution density in the Kelvin equation is incorrect. See Kreidenweis et al. (2005, ACP, pg. 1359). Discussion regarding the use of solution density should be removed. This is true as we are assuming an ideal aqueous phase. We have now corrected this in all the calculations and revised the manuscript accordingly.

Surface tension (Table 3, Eq. 1, pg. 28536): “calculating the surface tension as weight averages of the water and pure organic values”. I am unaware for the basis for that treatment. The surface tension of the aqueous solution usually follows a logarithmic law (e.g. the Szyzkowski equation) derived from Gibbs adsorption isotherms. Furthermore, application of surface tension values derived from bulk solutions produces incorrect results since the surfactant cannot lower the free energy of the surface and contribute to the water activity simultaneously. A correct theoretical treatment will have to include accurate partitioning between the phases (Sorjamaa et al., ACP, 2004). Since surfactant treatment is not at the center of this work, I recommend removing it completely and assuming the surface tension of pure water. This is a valid concern. The linear weighing for mixture surface tensions is typically applicable only for hydrocarbon mixtures or other mixtures with relatively similar components (see e.g. Poling et al., 2001), and indeed more sophisticated approaches for aqueous solutions are based on correction terms applied to the surface tension of pure water (see for example Werner et al. 2014 and references therein). We have chosen the linear weighing in an effort to introduce a simple approach for the lowering effect of the organics on the water surface tension – particularly taking into account the fact that we used only one representative value for the organic surface tension. On another note, the linear approach turns out to be a reasonable approximation for succinic acid, although it underestimates the impact of the acid on the mixture surface tension (see Werner et al. 2014 and references therein). However, to avoid this kind of complexity, we have redone all the calculations assuming the surface tension of water and have updated the figures in the revised manuscript accordingly.

Eq. (5): the units are a mix between mass and molar (csat is g/g and Yi is mole fraction? Maybe it is just a typo in the text below?). Equation 5 is correct, since analogously to the gas-liquid equilibrium the organic-aqueous phase equilibrium is dependent on the mole fraction in the insoluble organic phase, as long as the activity coefficient is also defined on a molar basis. In principle, similar formulation could be made with mass-based definitions. In our case, however, the two approaches are equivalent due to the assumption of constant molar mass throughout the organic mixture. We have clarified this in the revised manuscript.

The statement that “water solubility determines the aerosol water content at equilibrium” is wrong or only partially correct, depending on how one views the problem. The water content at equilibrium is controlled by RH, surface tension, and the molecular weight and activity of the dissolved components. If all components are dissolved, solubility exerts zero control on aerosol water content (as acknowledged in the paper). True. We have rephrased this sentence in the revised manuscript.

The distinction between ideal mixture vs unit activity model is unclear. In an ideal solution, the activity coefficient of each component is equal to unity. Thus, the two models are semantically identical. Despite several attempts I am unable to extract from the paper how the two cases are different. In the ideal mixture case, on the other hand, the mixture activity (the product of activity coefficient and mole
fraction) is unity – indicating that the individual components dissolve to water in the same way as they
would if they were present in their pure form. We have added this clarification to the revised manuscript.

(13) Comparing the limited solubility model to the full dissolution model is a straw man argument. Of
course they will agree if all species will dissolve, disagree if they do not, and the disagreement will be
proportional to ct. It is trivial that if one breaks the assumption of a model that it won’t work. I suggest
removing that discussion from the paper.

We understand the point that the Reviewer is making, but think there is indeed value in quantifying the
actual differences in activity diameter predictions – given that we have explored a large fraction of all
atmospherically relevant cases. This helps in putting the importance of solubility in context with other
potential uncertainties present in modeling the atmospheric cloud formation processes. What we also
think is remarkable in our results is that assuming a single ct for all the mixtures yields a reasonably
good prediction of the soluble fraction and thus the activation diameters.

(14) As far as I can tell, the epsilon model and the kappa model are identical since kappa =
epsilon*kappa_max and kappa_max is fixed in the study. Therefore figures 7a,b are redundant, as are
figures 11, e.e. There may be merit to keeping both treatments if molecular weight of the organic is
allowed to vary.

This is partly true, as κ or ε are directly related. However, the sensitivity of the activation diameter to
these parameters is different, since when the ε model is used, more information needs to be given in the
fitting process. However, the results shown in Fig. 6 of the discussion paper show that this information
does not make a big difference in the predicted activity diameters – although there is a small effect seen
in the narrower quartile range present for the ε model.

(15) Figure 7b suggests that there are numerical issues with the model. How can the fitted kappa exceed
kappa max?

This was due to using the solution density and surface tension in the Kelvin term of the full model
calculations while this information was not given to the simplified schemes. As suggested by the
reviewer, we have now repeated all the calculations assuming the surface tension and density of pure
water, and have modified the related figures accordingly.

(16) It would be helpful to include isolines of epsilon in Figure 3,b to visualize the range of solubilities
probed by the model.

This is a good suggestion and we have redrawn the corresponding figure adding the suggested isolines.

(17) Page 28534, line 5 to end of paragraph: Is it necessary to list how many Köhler curves were made?
More interesting would be e.g. calculations of d(property)/d(other property) in your calculations.

Indeed, the total number of the Köhler curves is perhaps not that interesting, but reason we have included
the number of the activation points just to illustrate the number of different combinations of solubility
distributions, particle sizes and supersaturations we investigated. We are not sure what the reviewer
means by calculations of d(property)/d(other property), and would prefer to keep the description as it is.

(18) Page 28525, Typo line 11, “Aalbrecht”

We have corrected the typo.

(19) Page 28525, Line 19-25: connect with ‘and’ ?

We have corrected the typo.
We have modified the revised manuscript accordingly.

We have clarified this in the revised manuscript.

We have expanded the paragraph as suggested. However, without further information it is difficult to know what oversimplification the Reviewer refers to, and would prefer to keep the table as it is.

We have revised the title to read "Equilibrium between water vapour and an aqueous phase containing dissolved material"

Corrected.

Corrected.

References


Connecting the solubility and CCN activation of complex organic aerosols: A theoretical study using solubility distributions

I. Riipinen¹,², N. Rastak¹ and S. N. Pandis²,³

¹ Department of Environmental Science and Analytical Chemistry, Stockholm University, Stockholm, Sweden
² Center of Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA, United States
³ Department of Chemical Engineering, University of Patras, Patras, Greece

Correspondence to: I. Riipinen (ilona.riipinen@aces.su.se)

Abstract

We present a theoretical study investigating the cloud activation of multicomponent organic particles. We modeled these complex mixtures using solubility distributions (analogous to volatility distributions in the VBS, i.e. volatility basis set, approach), describing the mixture as a set of surrogate compounds with varying water-solubilities in a given range. We conducted Köhler theory calculations for 144 different mixtures with varying solubility range, number of components, assumption about the organic mixture thermodynamics and the shape of the solubility distribution, yielding approximately 6000 unique cloud condensation nucleus (CCN)-activation points. The results from these comprehensive calculations were compared to three simplifying assumptions about organic aerosol solubility: 1) complete dissolution at the point of activation; 2) combining the aerosol solubility with the molar mass and density into a single effective hygroscopicity parameter κ; 3) assuming a fixed water-soluble fraction εeff. The complete dissolution was able to reproduce the activation points with a reasonable accuracy only when the majority (70-80%) of the material was dissolved at the point of activation. The single parameter representations of complex mixture solubility were confirmed to be powerful semi-empirical tools for representing the CCN activation of organic aerosol, predicting the activation diameter within 10% in most of the studied supersaturations. Depending mostly on
the condensed-phase interactions between the organic molecules, material with solubilities larger than about 0.1-100 g L\(^{-1}\) could be treated as soluble in the CCN activation process over atmospherically relevant particle dry diameters and supersaturations. Our results indicate that understanding the details of the solubility distribution in the range of 0.1 to 100 g L\(^{-1}\) is thus critical for capturing the CCN activation, while resolution outside this solubility range will probably not add much information except in some special cases. The connections of these results to the previous observations of the CCN activation and the molecular properties of complex organic mixture aerosols are discussed. The presented results help unravel the mechanistic reasons behind observations of hygroscopic growth and CCN activation of atmospheric secondary organic aerosol (SOA) particles. The proposed solubility distribution framework is a promising tool for modeling the interlinkages between atmospheric aging, volatility and water-uptake of atmospheric organic aerosol.

1 Introduction

Interactions of atmospheric aerosol particles with ambient water vapour determine to a large extent the influence that aerosols have on climate. On one hand the water content of aerosol particles at atmospheric relative humidity (RH) below 100% contributes significantly to the direct effect they have on the global radiative balance (Seinfeld and Pandis, 2006; Petters and Kreidenweis, 2007; Swietlicki et al., 2008; Zieger et al., 2011; Rastak et al., 2014). On the other hand the water-affinity of aerosol constituents, together with their dry size, defines the efficiency with which these particles can activate as cloud condensation nuclei (CCN) at supersaturated conditions (RH > 100%), form cloud droplets, and thus affect the properties of clouds (Twomey 1974; Aalbrecht, 1989; McFiggans et al., 2006). To quantify the effects of aerosol particles on clouds and climate it is thus necessary to understand the ways that aerosol constituents interact with water.

Organic compounds contribute a large fraction (20-90%, depending on the environment) of atmospheric submicron particulate mass (Jimenez et al., 2009) – which is the part of the aerosol size distribution that typically dominates the CCN numbers. A significant fraction of this organic aerosol (OA) is secondary – i.e. produced in the atmosphere from the condensation of oxidation products of volatile, intermediate volatility and semi-volatile organic compounds (VOCs, IVOCs and SVOCs). Emissions of biogenic VOCs such as monoterpenes, isoprene and
sesquiterpenes, followed by their subsequent oxidation and condensation in the atmosphere are thought to be the dominant source of secondary organic aerosol (SOA) on a global scale (Hallquist et al., 2009 and references therein) – although recent studies suggest also a notable anthropogenic component to the global SOA (Volkamer et al., 2006; Hoyle et al., 2011; Spracklen et al., 2011).

The solubility in water is one of the key properties governing the water-absorption (i.e. hygroscopic growth) and CCN activation of aerosol particles. Together with aqueous phase activity coefficients, surface tension, density and dry mass of the particle, water-solubility affects the aerosol particle water content in thermodynamic equilibrium (Pruppacher and Klett, 1997; Seinfeld and Pandis, 2006; Topping et al., 2012). Atmospheric organic compounds have a wide range of solubilities (Raymond and Pandis, 2003; Chan et al., 2008; Psichoudaki and Pandis, 2013). OA is thus a complex mixture of molecules with different CCN-behaviour than pure compounds. To accurately predict the water content and CCN activation of atmospheric OA, information on the dissolution behaviour and aqueous phase interactions of these complex mixtures is needed.

Representation of the complexity of OA is a major challenge for atmospheric chemical transport models: OA consists of thousands of different compounds whose properties are poorly known (Golstein and Galbally, 2007; Hallquist et al., 2009; Kroll et al., 2011). Approaches that simplify the complex nature of the OA mixture, yet reproducing its behaviour accurately enough, are required to be able to assess the climate and air quality effects of atmospheric organics in large-scale modelling applications. One example of such approach is the representation of the condensation and evaporation of SOA using a limited number of surrogate compounds with a range of saturation concentrations, known as the volatility basis set (VBS, Donahue et al., 2006; 2011; 2012). Similar simplifying approaches are needed to represent the hygroscopic growth and CCN activation of OA as well.

When interpreting laboratory and field studies on hygroscopicity and CCN activation, a number of simplifying assumptions about the OA properties have been made, for instance: 1) assuming that organics completely dissolve in water at the point of activation (Huff-Hartz et al., 2006); 2) assuming a fraction ($\varepsilon_{\text{eff}}$) of organics to be completely soluble and the remaining fraction ($1 - \varepsilon_{\text{eff}}$) completely insoluble in water (e.g. Pruppacher and Klett, 1997; Engelhart et al., 2008); 3) lumping the phase-equilibrium thermodynamics, molar masses and densities of the OA constituents into a single semi-empirical parameter. One of the most commonly-used
formulations is the hygroscopicity parameter $\kappa$, which relates the water activity in the aqueous solution with the water and dry particle volumes, and can be modified to account for limited solubility as well if the solubilities of the individual aerosol constituents are known (Petters and Kreidenweis, 2007; 2008; 2012, Petters et al., 2009a-c; Farmer et al., 2015; see also Rissler et al., 2004 and Wex et al., 2007 for alternative single-parameter formulations). These common simplifications of organic aerosol solubility and hygroscopicity are summarized in Table 1.

Laboratory studies on different types of organic aerosols have provided important insights into the relationship between CCN activation, hygroscopic growth and water-solubility of the atmospheric OA constituents. Raymond and Pandis (2002, 2003) and Chan et al. (2008) investigated the CCN activation of single- and multi-component aerosol particles consisting of organic compounds with known solubilities in water, and found that the particles activated at lower supersaturations than would have been expected based on the bulk solubility of their constituents. As an example, the laboratory studies by Chan et al. (2008) indicate that the CCN activation of material with water solubility as low as 1 g L$^{-1}$ could be predicted assuming complete dissolution. For some model systems the surface properties (wettability) of the aerosol particles, instead of the bulk water-solubility, seemed a more important factor defining their CCN activation (Raymond and Pandis, 2002). Huff Hartz et al. (2006) assigned part of this effect to residual water left in the particles upon their generation, causing the particles to exist as metastable aqueous solutions and thus activate at lower supersaturations than the corresponding dry material. The rest of the apparent increase in solubility was assigned to potential impurities in the particles. In general the results reported by Huff Hartz et al. (2006) suggested that compounds with water-solubilities above 3 g L$^{-1}$ behaved as if they were completely soluble in water, in general agreement with the earlier results of Hori et al. (2003).

Secondary organic aerosol particles generated in the laboratory through oxidation chemistry and condensation of the reaction products have also been found to activate as cloud droplets and thus contribute to the atmospheric CCN budgets (Cruz and Pandis, 1997; 1998; Huff Hartz et al., 2005; VanReken et al., 2005; Prenni et al., 2007; King et al., 2007; 2009; Engelhart et al., 2008; 2011; Asa-Awuku, 2009; 2010). These particles probably resemble the real atmospheric SOA more closely than individual organic species or their simple mixtures, but the theoretical interpretation of their CCN-behaviour is complicated by the variety of their constituents. Despite the fact that CCN-activity of SOA has been reported to vary with the volatile precursor identity and loading (Varutbangkul et al., 2006; King et al., 2009; Good et
al., 2010), photochemical aging (Duplissy et al., 2008; Massoli et al., 2010, and temperature (Asa-Awuku et al., 2009), the reported hygroscopicity parameter $\kappa$ values determined for different SOA types are remarkably similar, being typically around 0.1 for the overall SOA and 0.3 for the dissolved fraction extracted from the aqueous sample (Asa-Awuku et al., 2010; King et al., 2010). Similarly, Huff Hartz et al. (2005) reported effective solubilities of as high as 100 g L$^{-1}$ for both mono- and sesquiterpene SOA – although both are known to consist of a range of compounds with different solubilities. These results demonstrate the importance of knowing the water-soluble fraction of SOA at varying conditions but suggest that its exact speciation is probably not necessary for predictive understanding of the CCN activity of SOA particles (Asa-Awuku et al., 2010; Engelhart et al., 2011). The $\kappa$ values inferred from sub- or supersaturated conditions for the same SOA mixtures, on the other hand, are not always consistent, the sub-saturated $\kappa$ values being typically lower than the super-saturated ones (Prenni et al., 2007; Duplissy et al., 2008; Wex et al., 2009; Topping et al., 2012). Multiple possible reasons for this have been presented in the literature, including incomplete dissolution of the aerosol constituents at sub-saturated conditions (Petters et al., 2009a), surface tension effects (Good et al., 2010), RH-driven effects on the reaction chemistry and thus the composition (solubility and activity) of the formed SOA (Poulain et al., 2010), evaporation and condensation of semi-volatile organic compounds (Topping et al., 2012), or the non-ideality of the mixtures being more pronounced at the sub-saturated conditions (Kreidenweis et al., 2006; Petters et al., 2009a; Good et al., 2010).

While the basic theory of cloud droplet activation for pure water-soluble compounds is relatively well-established (Pruppacher and Klett, 1997; Asa-Awuku et al., 2007; Topping et al., 2012; Farmer et al., 2015), and a number of theoretical and experimental studies on the different aspects controlling the CCN-activation of SOA have been presented (see above, McFiggans et al., 2006; Dusek et al., 2006 and references therein), only few of these studies have investigated the implications of the water-solubilities of complex organic mixtures on CCN activation. Understanding the relationship between the dissolution behaviour and CCN-activation of complex organic mixtures is, however, needed to constrain the water-soluble fraction of SOA in varying conditions as well as to systematically unravel the mechanisms causing the apparent simplicity in the CCN-behaviour of complex organic mixtures.

In this work we introduce a framework for representing the mixture components with a continuous distribution of solubilities, similar to the VBS (Donahue et al., 2006; 2011; 2012).
Using this framework in a theoretical model, we investigate the dissolution behaviour of complex organic mixtures and their CCN activity, focusing on the impact of mixture solubility on CCN-activation. In particular, we study the response of the CCN-activation to varying solubility ranges, distribution shapes, and numbers of components in the mixture. Furthermore, we compare the CCN-activation predictions using the simplified solubility representations outlined above (complete dissolution, soluble fraction $\varepsilon_{\text{eff}}$, and hygroscopicity parameter $\kappa$ without including knowledge about the component solubilities) with the more detailed description using the full solubility distributions, and study the relationship of the simplified solubility parameters $\varepsilon_{\text{eff}}$ and $\kappa$ to the true mixture solubility distribution. Although the solubility ranges and other thermodynamic properties of the mixture have been chosen to represent SOA, many of the concepts and approaches introduced here can be applied to any particles consisting of complex mixtures of organic compounds with varying water-solubilities. Finally, we discuss the applicability of the introduced framework for describing the water-interactions of realistic SOA mixtures and the relevant future directions.

2 Methods

2.1 Theoretical predictions of CCN activation of complex organic mixtures

Figure 1 schematically summarizes the model system considered in this study. We consider a monodisperse population of spherical aerosol particles consisting of an internal mixture of organic compounds. When exposed to water vapour, these particles grow reaching thermodynamic equilibrium between the water vapour and the particle phase. The wet particle is allowed to consist of maximum two phases: the insoluble organic phase and the aqueous phase. The compositions of the organic and aqueous phases are determined on the one hand by the equilibrium between the aqueous phase and the water vapour, and on the other hand by the equilibrium of the aqueous phase with the organic insoluble phase. To isolate the effects of solubility from organic volatility effects, we do not allow the organics to evaporate from the droplet – i.e. we assume that the equilibrium vapour pressures of the organics is zero above the droplet surface. Similarly, no condensation of organics from the gas phase to the particles is allowed to take place. The validity of this assumption depends on the gas-phase concentrations of the organic species as well as the atmospheric temperature during the cloud formation process. Testing it at different atmospherically relevant conditions deserves some future
attention, accounting for the dynamics of the atmospheric gas-phase as well (see also Topping et al., 2012). In this study, however, we focus strictly on the CCN-activation process. The organic composition and dry particle size were treated as an input to a model calculating the final equilibrium composition, wet size, and CCN activation behaviour of these particles. Note that while the solubility in the equations presented in the next Sections 2.1 and 2.2 is non-dimensional (g g\textsubscript{H2O}^{-1}), in the presentation of the results it is converted into g L^{-1}, assuming constant unit density of water.

2.1.1 Equilibrium between water vapour and an aqueous phase containing dissolved material

The Köhler equation (Pruppacher and Klett, 1997) is used to link the ambient water vapour saturation ratio $S$ with the size, composition and water content of the aerosol particles in thermodynamic equilibrium (lower panel of Fig. 1):

$$S = \frac{p_{w,eq}}{p_{w,sat}} = a_w \exp\left(\frac{4\sigma v_w}{RTD_{p,wet}}\right),$$

where $p_{w,eq}$ (Pa) is the equilibrium vapor pressure of water over the droplet surface, $p_{w,sat}$ (Pa) the saturation vapor pressure over a pure flat water surface, $\sigma$ (N m\textsuperscript{-1}) is the surface tension of the droplet, $v_w$ the molar volume of water in the aqueous phase, $M_w$ (kg mol\textsuperscript{-1}) the molar mass of water, $\rho$ (kg m\textsuperscript{-3}) the density of the aqueous phase, $D_{p,wet}$ (m) the diameter of the droplet, $T$ (K) the temperature and $R$ (J mol\textsuperscript{-1} K\textsuperscript{-1}) the universal gas constant. $a_w$ is the water activity, defined as the product of the water mole fraction $X_w$ and water activity coefficient in the aqueous phase $\Gamma_w$:

$$a_w = X_w \Gamma_w.$$  

The activity coefficient describes the interactions between water molecules and the dissolved organic molecules in the mixture. The saturation ratio at which the particles of dry size $D_{p,dry}$ activate as cloud droplets (i.e. continue growing in size even if the saturation ratio decreases), is referred to as the critical saturation ratio $S_c$. Mathematically this corresponds to the highest local maximum in the $S(D_{p,wet})$ curve, usually referred to as the Köhler curve.

2.1.2 Equilibrium between the aqueous and insoluble organic phases

The composition of the droplet and the distribution of material between the organic insoluble and the aqueous phases can be calculated applying the principles of mass conservation and the
thermodynamic equilibrium of the organic components in an aqueous mixture with the
insoluble organic phase. As the mass transfer of organics between the particles and the gas
phase is neglected, the total mass of the dry particle $m_{\text{dry}}$, being the sum over all components $i$, is equal to the total organic mass in the wet droplet (see Fig. 1):

$$m_{\text{dry}} = \sum_{i} m_{i,\text{insoluble}} + \sum_{i} m_{i,\text{aqueous}},$$

(3)

where $n$ is the total number of organic compounds, $m_{i,\text{insoluble}}$ is the mass of compound $i$ in the insoluble organic phase and $m_{i,\text{aqueous}}$ the mass of compound $i$ in the aqueous phase. The same holds for each organic compound individually:

$$m_{i,\text{dry}} = y_{i,\text{dry}} m_{\text{dry}} = m_{i,\text{insoluble}} + m_{i,\text{aqueous}},$$

(4)

where $y_{i,\text{dry}}$ is the mass fraction of $i$ in the dry organic particle. On the other hand, the concentration of each organic compound in the aqueous phase is determined by the thermodynamics of the two-phase system consisting of the insoluble organic phase and the aqueous solution phase. The mass of each organic compound $i$ in the aqueous phase can be expressed as (Prausnitz et al., 1998; Banerjee, 1984):

$$m_{i,\text{aqueous}} = \begin{cases} 0, & Y_{i,wet} > 0 \\ \frac{Y_{i,wet} c_{\text{sat, pure},i} m_{w}}{m_{i,\text{dry}}}, & Y_{i,wet} = 0 \end{cases},$$

(5)

where $Y_i$ is the activity coefficient of $i$ in the insoluble organic phase (where the reference state is the pure component dissolution to water), $Y_{i,wet}$ and $c_{\text{sat, pure},i}$ (here in g g$^{-1}$) are the organic phase mole fraction and pure component solubility (saturation concentration) of $i$, and $m_w$ is the total mass of water in the droplet. The former equation corresponds to the situation where the particle contains an insoluble organic core in thermodynamic equilibrium, the latter to the case where only the aqueous phase exists, i.e. all the organic material has dissolved to the water. Although the mole fraction and the corresponding molar activity coefficient have been used in Eq. 5, a similar relationship can be defined using the mass fraction in the organic phase and a corresponding mass-based activity coefficient. For a multicomponent system in which the molar mass of the organic species varies, the mole and mass fractions of a given species are not necessarily equal. In this study, however, we assume a constant molar mass throughout the organic mixture for simplicity, leading to the mass and mole fractions in the organic phase to be the same, i.e. $Y_i = y_i$ for all compounds. All the equations presented below can be re-derived
in a relatively straightforward manner taking into account a potential difference between the mole and the mass fractions in the organic phase.

Finding the organic and aqueous phase compositions that satisfy Eqs. 3-5 for given water and dry particles masses \((m_w, m_{dry})\), respectively, requires solving \(n\) coupled equations. These equations were expressed using the ratio \(\chi_i\) of organic compound \(i\) in the insoluble core of the wet particle to the total mass of the compound (Raymond and Pandis, 2003; Petters and Kreidenweis 2008):

\[
\chi_i = \frac{m_{i, insoluble}}{m_{i, insoluble} + m_{i, aqueous}} = \frac{m_{i, insoluble}}{m_{i, dry}} \cdot \frac{Y_{i, dry} m_{dry}}{m_{Y}}. \tag{6}
\]

The mole fraction (equal to the mass fraction for the mixtures considered here) of \(i\) in the insoluble core is defined as

\[
\chi_i = \frac{\sum_i m_{i, insoluble}}{\sum_i m_{i, dry}} = \frac{\sum_i \chi_i m_{i, dry}}{\sum_i \chi_i Y_{i, dry}}. \tag{7}
\]

Finally, combining equations 3-7, we get \(n\) equations of the form

\[
\chi_i = 1 - \frac{\chi_i Y_{i, dry} c_{i, sat, pure} m_w}{m_{i, dry} \sum_i \chi_i Y_{i, dry}}, \tag{8}
\]

which can be solved for \(\chi_i\) with the constraint \(0 \leq \chi_i \leq 1\) for given water and dry particle masses.

**2.1.3 Representation of complex organic mixtures: Solubility distributions and thermodynamic properties**

A novel aspect of this study as compared with previous theoretical work is the representation of complex mixtures using their aqueous solubility distribution of the individual species. In our calculations we used mixtures of \(n\) compounds, whose water-solubilities ranged from \(c_{sat,min}\) to \(c_{sat,max}\), either on a linear or logarithmic basis. The shape of the distribution could vary as well. In this work we studied essentially three types of mass fraction distributions in the dry particle: a uniform distribution in which all solubilities are equally abundant, distribution increasing steadily (linearly or logarithmically), and a distribution decreasing steadily (linearly or logarithmically). The 72 studied solubility distributions are specified in Table 2, and the solubility distributions for \(n = 10\), \(c_{sat,min} = 0.1 \text{ g L}^{-1}\) and \(c_{sat,max} = 1000 \text{ g L}^{-1}\) are presented in Fig. 2 as examples.
For simplicity, we assumed that water forms an ideal solution with the dissolved organics, i.e. \( \Gamma_w = 1 \), thus yielding an activity equal to the mole fraction of water, \( a_w = X_w \) in Eq. 1. Since information about the activity coefficients of organic species in purely organic mixtures is still scarce, we studied two alternative approaches to represent the dissolution thermodynamics of the SOA mixture in Eqs. 9-12: 1) assuming an ideal organic mixture where \( \gamma = 1 \) for all compounds in the insoluble phase; 2) assuming a constant organic phase activity \( \gamma Y_{i,wet} \) of unity for all compounds – in which case the dissolution behaviour of each \( i \) is similar to their behaviour as pure components. These cases probably represent the limiting cases for the dissolution of SOA components in CCN activation reasonably well, the former representing a lower limit and the latter an upper limit for the overall solubility of the dry particle. Applying the two limiting assumptions about the interactions of the compounds in the organic phase for the 72 different solubility distributions (Table 2) thus results in total 144 unique representative model mixtures.

The density, surface tension, and molar masses assumed for water and the organic compounds are summarized in Table 3. Although the density, surface tension and molar mass of the organics are likely to vary with the solubility, we kept them constant throughout the organic mixture to isolate the solubility effects on the CCN behaviour. The values were chosen based on literature studies of the CCN behaviour of SOA (Engelhart et al., 2008; Asa-Awuku et al., 2010). The surface tension \( \sigma \) was approximated by the surface tension of water, and the molar volume of water in the aqueous phase was assumed the same as for pure water. Furthermore, we assumed no dissociation of the organics in the aqueous phase.

### 2.1.4 Model calculations

We solved Eq. 8 for organic mixtures with the Matlab internal function `fsolve`, for varying water and dry particle masses \( m_w \) and \( m_{dry} \), covering 50 different dry particle diameters between 20 and 500 nm. The calculations yielded the composition of the insoluble organic and the aqueous phase, and thus the mole fraction of water in the aqueous solution \( X_w \). From these results the Köhler curves \( S(D_{p,wet}) \) corresponding to each dry particle mass could be calculated using Eq. 1 (see Fig. 3a for an example of the Köhler curves). The critical supersaturations \( s_c \) (defined as \( S_c - 1 \)) corresponding to specific dry particle diameters \( D_{p,dry} \) (termed also as activation diameters \( D_{p,act} \) at a given saturation ratio \( S \) or supersaturation \( s \)) were determined from the maxima of the Köhler curves (see Fig. 3a). The temperature was assumed to be 298 K in all calculations. These calculations for the 144 unique organic model mixtures corresponded to
7200 Köhler curves yielding 5957 \((D_{p,act}, s_c)\) pairs (activation points, see Fig. 3b). For the remaining 1143 curves no activation points were found with the given combinations of mixture properties and dry diameters. For comparisons with the simple solubility representations, the dissolved organic fraction defined as

\[
\varepsilon = \sum \frac{m_{i,aqueous}}{m_{dry}}
\]  

was extracted from the model output.

### 2.2 Comparison of the full model output to simple solubility representations

To investigate the performance of the simple solubility representations given in Table 1 in reproducing the CCN activation of complex mixtures, we fitted the \((D_{p,act}, s_c)\) data created by the full model using these simpler models. No fitting is required for the complete solubility approach. Using the obtained solubility parameters from the optimal fit and the corresponding simplified forms of the Köhler equation we then recalculated new \((D_{p,act}, s_c)\) pairs and compared them to the predictions by the full model. Furthermore, we investigated the relationships between the true mixture solubility distribution and the simplified solubility parameters. The details of the approach used for each simple model are outlined below.

#### 2.2.1 Complete dissolution

In the case where all of the organic material is assumed to completely dissolve at the point of activation, the calculation of the aqueous solution composition becomes trivial as

\[
m_{i,dry} = m_{i,aqueous}
\]  

for all the compounds, and the water mole fraction can simply be calculated based on the dry particle mass as

\[
X_w = \frac{m_w}{M_w + \frac{m_{dry}}{M_{org}}},
\]

where \(m_w\) is the water mass in the droplet, \(m_{dry}\) the dry particle mass (related to \(D_{p, dry}\) through the organic density \(\rho_{org}\)) and \(M_{org}\) the organic molar mass. The \(X_w\) calculated in this way was inserted into Eq. 1 to yield the corresponding \((D_{p, act}, s_c)\) predictions and was also applied for
calculating the solution density and surface tension as mass-weighted averages of the water and pure organic values.

### 2.2.2 Hygroscopicity parameter $\kappa$

In many practical applications the water activity and the difference in the densities and molar masses of water and the dry material are expressed with a single hygroscopicity parameter $\kappa$, introduced by Petters and Kreidenweis (2007), defined as:

$$\frac{1}{a_w} = 1 + \kappa \frac{V_s}{V_w},$$  

where $V_s$ and $V_w$ are the volumes of the dry material and water, respectively. The following formulation of the relationship between water saturation ratio, aerosol size and composition is referred to as the $\kappa$-Köhler equation:

$$S = \frac{D_{p,act}^3 - D_{p,dry}^3}{D_{p,dry}^3 - D_{p,dry}^3 (1 - \kappa)} \exp \left( \frac{4\sigma M_w}{RT \rho D_{p,wet}} \right)$$  

yielding an approximate expression for the relationship between $s_c$ and $D_{p,act}$ defined as

$$s_c = \frac{2}{3} \left( \frac{4M_w \sigma}{RT \rho} \right)^{\frac{1}{2}} \left( 3\kappa D_{p,act}^3 \right)^{\frac{1}{2}},$$  

Equation (14) was fitted to all ($D_{p,act}$, $s_c$) data produced for a given organic mixture composition (see Table 2) by the full model, thus assuming a constant $\kappa$ value for a given organic mixture. To mimic the application of Eq. 14 to experimental data with no knowledge on the exact solute composition, in this case we assumed the surface tension and density to be those of water when fitting the $\kappa$ values to the full model data. The above formulation of $\kappa$, which is often used in the interpretation of experimental data as well, thus contains information about solubility, potential aqueous-phase non-ideality, as well as molar mass and density of the solutes (see Farmer et al., 2015).

### 2.2.3 Soluble fraction $\varepsilon_{eff}$

For an ideal solution of water and an organic solute the $\kappa$ is directly proportional to the dissolved fraction and the ratio of the molar volumes of water and the solute i.e.: $\kappa = \varepsilon \kappa_{max}$, where $\kappa_{max} = (M_w/M_{org})(\rho_{org}/\rho_w)$. Assuming that a single soluble fraction $\varepsilon_{eff}$ can represent a given organic
mixture (see Table 2) at all considered supersaturations, and substituting these relationships into Eq. 16 yields
\[
s_c = \frac{2}{3} \left( \frac{4M_w \sigma}{RT \rho} \right)^{\frac{3}{2}} \left( 3 \frac{M_w}{M} \rho_s \varepsilon_{p,act} D_p^{3} \right)^{-\frac{3}{2}},
\] (15)
the corresponding form of the Köhler equation being (see Huff Hartz et al., 2005)
\[
S = \frac{M_{org} \rho_w (D_{p,wet}^3 - D_{p,act}^3)}{D_{p,wet}^3 (M_{org} \rho_w) + D_{p,act}^3 (\varepsilon_{eff} M_w \rho_{org} - M_{org} \rho_w)} \exp \left( \frac{4 \sigma M_w}{RT D_{p,wet}} \right).
\] (16)
Again, we fitted Eq. 17 to the data produced by the full model and assumed the aqueous solution density and surface tension to be equal to those of water. When \( \varepsilon_{eff} < 1 \), the following relationship has been used to estimate the effective saturation concentration of the mixture (Raymond and Pandis, 2002; Huff Hartz et al., 2005)
\[
\varepsilon_{sat,eff} = \rho_w (D_{p,wet}^3 - D_{p,act}^3)
\] (17)
\[
\varepsilon_{eff} = \frac{\sum_{i=1}^{i} m_{i,aqueous} + \sum_{j=i+1}^{a} m_{j,aqueous}}{m_{dry}} = \frac{\sum_{i=1}^{i} m_{i,aqueous} + (\sum_{j=i+1}^{a} m_{j,dry} - \sum_{j=i+1}^{a} m_{j,insoluble})}{m_{dry}}.
\] (18)
We now hypothesize that assuming a single soluble fraction for a given aerosol mixture is in fact equivalent to assuming that everything above \( i \) is completely dissolved while all the material below this threshold remain undissolved, i.e.
\[
\varepsilon = \varepsilon_{sat,eff} = \frac{\sum_{j=i+1}^{a} m_{j,dry}}{m_{dry}}.
\] (19)
On the other hand, \( \varepsilon = \varepsilon_{eff} \) if the following condition is fulfilled (see Eq. 20):
Substituting Eq. 5 to Eq. 20 we now have

\[
\sum_{i=1}^{n} m_{i, \text{aqueous}} = \sum_{j=1}^{n} m_{j, \text{insoluble}} = \sum_{j=1}^{n} m_{j, \text{dry}} - \sum_{j=1}^{n} m_{j, \text{aqueous}}. \tag{20}
\]

At the limit of large \( n \) and in the case of a symmetric distribution of material between the insoluble organic and aqueous phases, Eq. 21 is satisfied by setting the threshold solubility \( i_t \) so that

\[
\lim_{i \to i_t} F_w \gamma_j Y_j, \text{sat, pure}_i, = Y_{\text{dry}, i_t} - F_w \gamma_j Y_j, \text{sat, pure}_i. \tag{23}
\]

In this case the threshold solubility \( c_t \) is found from the bin for which

\[
c_{\text{sat, pure}, i_t} = c_t \approx \frac{Y_{\text{dry}, i_t}}{Y_{\text{wet}, i_t}} \cdot \frac{1}{2 F_w}. \tag{24}
\]

This is also equal to the bin where 50% of the material is partitioned in the insoluble phase, i.e.

\[
\chi_{i_t} = \frac{m_{i_t, \text{insoluble}}}{m_{i_t, \text{insoluble}} + m_{i_t, \text{aqueous}}} = \frac{1}{1 + \frac{m_{i_t, \text{aqueous}}}{m_{i_t, \text{insoluble}}}} = \frac{1}{2}. \tag{25}
\]

Finding the solubility threshold \( c_t \) requires knowledge on the ratio \( F_w \) (Eq. 22). \( F_w \) on the other hand, depends on the ambient supersaturation and the total soluble mass – thus introducing a supersaturation-dependence to the \( \epsilon \) given by Eq. 18 as well. The magnitude of \( F_w \) as a function of supersaturation can be estimated by substituting Eqs. 15 and 17 into the definition of \( F_w \) (Eq. 22) and after some rearranging, yields:

\[
F_w = \frac{\rho_w}{\rho_{\text{org}} \left( \frac{2 \epsilon}{s_c} \cdot \rho_{\text{org}} \cdot \frac{M_Y}{s_{\text{org}}} - 1 \right)} = \frac{\rho_w}{\rho_{\text{org}} \left( \frac{2 \epsilon \kappa_{\text{max}}}{s_c} - 1 \right)}. \tag{26}
\]
3 Results

Figure 3a displays examples of the Köhler curves obtained from solving Eqs. 1 and 6 for distribution 1 (the flat logarithmic distribution) with varying solubilities and \( n = 5 \), assuming that the organics form an ideal mixture with each other. Each curve corresponds to a different dry size, and the dots indicate the activation point (\( S_{crit} \) corresponding to the activation dry diameter \( D_{p,act} \)). Black dots indicate incomplete dissolution (\( \varepsilon < 0.99 \)) at the point of activation while red dots indicate that in practice all the organics are dissolved into the aqueous phase at the point of activation (\( \varepsilon > 0.99 \)). Qualitatively similar behaviour was observed for all the considered distributions: as the overall solubility of the mixture increases, the dissolution of the compounds increases, leading eventually to complete dissolution at the point of activation. The transition from a regime with two phases (aqueous + insoluble) to a single aqueous phase is visible in the two maxima in Fig. 3a, in accordance with Shulman et al. (1996) and Petters and Kreidenweis (2008).

Figure 3b illustrates the parameter space probed in this study, showing the \( (s_{crit}, D_{p,act}) \) points corresponding to the Köhler curve maxima calculated for all the considered organic mixtures. The relationships between the critical supersaturation, activation diameter, and dissolved fraction \( \varepsilon \) at the point of activation are also schematically shown. The chosen dry diameters and supersaturations represent a conservative range of typical atmospheric conditions – as the total aerosol number concentrations are dominated by ultrafine (diameters smaller than 100 nm) particles at most locations. In most considered cases the dissolved fractions fall between 0.1 and 1, but the lowest dissolved fractions at the point of activation are of the order of only a few percent – thus mimicking nearly insoluble aerosols. Therefore the cases considered here represent a reasonable sample of atmospherically relevant conditions and SOA mixture compositions. The water-to-organic mass ratios \( F_w \) corresponding to the probed conditions and mixtures (see Sect. 2.2.4) range from values below 1 up to 1000, with most values around 10-100. In many of the following plots and considerations we have chosen four specific supersaturations, 0.1%, 0.3%, 0.6% and 1%, as representative values for typical laboratory experiments, which are also indicated in Fig. 3b.

An example of the dependence of the activation diameter \( D_{p,act} \) on the solubility range for all the studied distributions (see Table 2) and \( n = 5 \) is presented in Fig. 4. As expected, the activation diameter decreases with increasing supersaturation and solubility range for a given solubility distribution. The solubility distribution is reflected in the overall magnitude of the
activation diameters: the distributions that have larger fractions of material in the higher end of
the solubility range (distributions 2, 3, and 4) have generally lower activation diameters for a
given supersaturation as compared with the other distributions. The case when unity activity in
the organic phase is assumed results in smaller activation diameters for the same supersaturation
as compared with the ideal organic mixture case (see Sect. 2.1.3 for the definitions of the cases).

Figure 5 presents the activation diameters predicted using the simplified solubility
descriptions (Table 1) based on best fits to all available data as compared with the full
description of the solubility distributions (Table 2). The results clearly show, not surprisingly,
that assuming complete dissolution for all the mixtures consistently under-predicts the
activation diameters (Fig. 5a). Representing the dissolution behaviour with only one additional
parameter, i.e. the hygroscopicity parameter $\kappa$ (Eqs. 15-16) or the effective soluble fraction $\varepsilon_{\text{eff}}$
(Eqs. 17-18) improves the agreement between the activation diameters considerably (Figs. 5b
and 5c). Adding the knowledge about the molar mass and density of the organic mixture, which
is the only difference between using the $\varepsilon_{\text{eff}}$ instead of the single $\kappa$, adds only marginal
improvements in predicting the activation diameters for a given supersaturation. The
disagreements between the simplified models and the full theoretical treatment are largest for
the smallest supersaturations. These are the cases with the widest range of possible $\varepsilon$ values at
the point of activation (see Fig. 3b), and the effect is most obvious for the complete dissolution
model: the larger the deviation from complete dissolution at the point of activation (i.e. the $\varepsilon \sim$
1 case in Fig. 3b) the more significant error we introduce. The activation diameters predicted
assuming complete dissolution are within 10% of the correct values if the real dissolved fraction
$\varepsilon$ is larger than about 0.7-0.8 at the point of activation.

The performance of the simple solubility models for all the studied Köhler curves is
summarized in Fig. 6: while the complete dissolution assumption results in systematic under-
prediction (up to 40%) of the activation diameter, the $\kappa$- and $\varepsilon_{\text{eff}}$-based models are generally
within 10% (in most cases within 5%) of the activation diameter predicted for the full solubility
distribution representation.

Figure 7 compares the fitted parameters representing the mixture dissolution to the
 corresponding values inferred from the full mixture data for the 144 different mixtures. In Fig. 7a, the effective soluble fractions $\varepsilon_{\text{eff}}$ are compared to the actual dissolved fractions $\varepsilon$ at the point
of activation for all the studied mixtures. While in the fits a single constant $\varepsilon_{\text{eff}}$ has been assumed
to represent a given mixture (see Eq. 15-16), in reality $\varepsilon$ varies with supersaturation (Eq. 9).
Thus, while the fitted $\varepsilon_{\text{eff}}$ for a given mixture correlates very well with the average $\varepsilon$ over all activation points (the markers in Fig. 7a), the performance of the approach can vary considerably with supersaturation (the grey lines in Fig. 7a). In practice this means that describing a given complex mixture with a fixed soluble fraction yields representative average dissolution behaviour, but does not guarantee correct solubility description for a specific $s_c$ if fitted over a range of supersaturations. The corresponding comparison for the hygroscopicity parameter $\kappa$ values describing the data are shown in Fig. 7b. A clear correlation between the fitted $\kappa$ and the average $\varepsilon$ is observed as expected (see Sect. 2.2.3), but the variation of $\varepsilon$ with supersaturation again adds scatter to the data – suggesting a dependence of $\kappa$ on $s_c$. The maximum $\kappa$, on the other hand, is defined primarily by the molar masses and densities of the organics. For our mixtures with constant $M_{\text{org}}$ and $\rho_{\text{org}}$ the value of $\kappa_{\text{max}}$ is 0.15, which is indicated in Fig. 7b. The points above this theoretical maximum are a result of using the pure water density instead of the mixture value in the Kelvin term of the Köhler equation (Eq. 16). These results thus suggest that the $\kappa$ values of 0.1-0.2 typically observed for SOA particles (Duplissy et al., 2011) are controlled by the molar masses and densities of the SOA mixtures to a large extent and can result from quite different SOA mixtures in terms of their solubilities.

To illustrate the relationship of the fitted $\varepsilon_{\text{eff}}$ to the dissolution of a given mixture, the partitioning between the aqueous and insoluble organic phase is presented in Fig. 8 for distribution 1 with the “low” solubility range and $n = 100$ at the point of activation when $s_c = 0.1\%$ (see Table 2). Figure 8a shows the partitioning for the case where ideal organic mixture has been assumed and Fig. 8b shows the corresponding data for the unity activity case (see Eq. 5). The point of 50%-partitioning ($c_t$, Eqs. 20-21) is also shown. As described in Sect. 2.2.4 we expect $c_t$ to be a reasonable estimate for the limit for complete dissolution, if the complex mixture is reduced into a two-component mixture of completely soluble and insoluble components. It should be noted, however, that the water content and $\varepsilon$ of the droplet at the point of activation depend on supersaturation (see Eq. 25), causing also a dependence between $c_t$ and $s_c$. Furthermore, Fig. 8 illustrates a difference in the solubility-dependence of the partitioning behavior for the two organic activity assumptions. The ideal mixture displays a symmetrical sigmoidal dependence around $c_t$. For the unity activity case, on the other hand, the undissolved fraction is asymmetric around the 50%-value – dropping rapidly to zero above $c_t$ but approaching 1 asymptotically below $c_t$. 
Figure 9 shows the distributions of the solubility bins containing the 50%-partitioning points ($c$, Eqs. 20-21) on a decadal basis for all the activation points studied, illustrating also the differences between the two assumptions about the organic phase activity. The $c$ values for the ideal organic mixture (Fig. 9a, based on 2465 points) display a symmetrical distribution around the median value of about 10 g L$^{-1}$. Also, a modest dependence of $c$ on the number of components is observed: the cases with 3 and 5 components display slightly higher $c$ values as compared with the cases with larger $n$. This apparent dependence is probably due to the discrete nature of the solubility distributions in combination with the fact that for the different solubility ranges (see Table 2) only the lower end of the distribution is changed while the upper end is always at 1000 g L$^{-1}$. The unity activity case displays a much stronger $n$-dependence (Fig. 9b, based on 3492 points): if analysed separately, the median $c$ shifts from about 0.1 to 10 g L$^{-1}$ when $n$ changes from 100 to 10 and 5, and up to 100 g L$^{-1}$ for $n = 3$. Unlike the ideal mixture this behavior is explained by the actual dissolution thermodynamics: in a system where the components do not affect each other’s solubility directly (i.e. the dissolution of a compound $i$ is only limited by its own presence in the aqueous phase), the amount of dissolved material is only dependent on the total water content and is larger the larger the number of dissolvable components. If all the different mixtures are integrated together the median $c$ for the unity activity assumption lies at about 1 g L$^{-1}$ – a decade lower than for the ideal mixture case. Figure 10 provides a more detailed look on the sensitivity of the $c$ to supersaturation, $n$ and molar mass for one of the distributions (Distribution 1, mid solubility range, see Table 2). As expected, $c$ depends considerably on supersaturation. It can also be seen that while $c$ shows some sensitivity to the number of components (in line with Fig. 9) and molar mass, in our case by far the most critical assumption is related to the organic mixture thermodynamics.

The fitted $\varepsilon_{\text{eff}}$ values are compared in Figure 11 to the fraction of mass with solubilities above the median $c$, for the 134 mixtures that activated at the probed conditions, individually for both organic activity assumptions. The fitted dissolved fraction corresponds well to the fraction of mass with solubilities above the 50%-partitioning point, as predicted by the theoretical principles outlined in Sect. 2.2.4. Also, the solubilities of the different distributions with varying shapes, numbers of components and solubility ranges can be represented reasonably well with a single median $c$ (equal to 10 g L$^{-1}$ for the ideal mixture case and 1 g L$^{-1}$ for the unity activity case, see Fig. 9) with median deviations between the fitted $\varepsilon_{\text{eff}}$ and the fraction above $c$ of 9% and 8%, respectively. On the other hand, these results indicate that the soluble fractions determined from experimental data on CCN activation provide information
about the fraction of material with solubility above \( c_t \). There are 8 points that do not seem to follow the general trend, however: a group of points with all the material above the threshold solubility can display a variety of fitted \( \varepsilon_{\text{eff}} \) values. These are all points that correspond to the “high” solubility ranges. Distributions 5 and 6 (see Table 2) with \( n = 3 \) are among these points for both organic activity assumptions. For the ideal mixture case also distribution 1 with 3 components diverges from the general trend, and for the unity activity assumption distribution 5 falls into the category regardless of the number of components. These points thus contribute to the high ends of the \( c_t \) distributions in Fig. 9.

The deviations in the activation diameters as predicted by the three simplified solubility representations (complete dissolution, \( \varepsilon \) and \( \kappa \)) are displayed in Figure 12 as a function of the mixture properties for \( s_c = 0.1\%, 0.3\%, 0.6\% \) and 1\%. Again, the two different assumptions about the organic phase activity are treated separately due to their different limiting solubilities \( c_t \) (Fig. 9) and different shapes of the partitioning distributions (Fig. 8). Also, the points close to complete dissolution at the point of activation (\( \varepsilon \geq 0.8 \), see Fig. 5 and its explanation in the text) are presented with a different color (gray symbols) than the points where the activation diameter differs significantly from the complete dissolution prediction (\( \varepsilon < 0.8 \), black symbols).

As expected, the complete dissolution assumption performs better for the more water soluble organic mixtures. Figs. 12a and 12b illustrate this by showing the relationship between the norm of the error in the predicted \( D_{p,act} \) and the fraction of material below the median \( c_t \) (10 and 1 \( g \) \( L^{-1} \)) for the two organic activity assumptions. The larger the amount of material below the solubility limit, the larger the deviation from the full model predictions. For the \( \kappa \) and \( \varepsilon \) models, on the other hand, the variable best correlating with the error in \( D_{p,act} \) induced by the simplification is different for the different organic activity assumptions – although close to complete dissolution these models also do well, nearly independent of the solubility of the distribution. For the ideal mixture case the fraction of mass between 1 and 100 \( g \) \( L^{-1} \) correlates better with the error (Fig. 12c, 12e) than the mass fraction below any solubility limit (not shown), while for the unity activity case the material at the low end of the distribution (mass fraction below 1 \( g \) \( L^{-1} \)) performs better (see Fig. 12d, f). The reason for this lies in the different shapes of the partitioning distributions resulting from the two assumptions (Fig. 8). For the symmetric partitioning curve of the ideal mixture case, the predicted \( \varepsilon \) and \( \kappa \) are most sensitive to differences in the partitioning behaviour between compounds within the range of \( c_t \) corresponding to the supersaturation and particle diameter ranges studied here, i.e. 1-100 \( g \) \( L^{-1} \) (see Figs. 12c, 12e). Anything outside these boundaries will behave as completely soluble or
insoluble throughout the studied supersaturation space, thus not introducing a significant error
when constant \( \varepsilon \) is assumed to describe the mixture. However, the more the material that can
behave as either insoluble or soluble depending on the conditions, the larger error we introduce
by assuming a constant \( \varepsilon \) for a given mixture at any conditions. The story is different for the
unity activity case (Figs. 12d, 12f): as the shape of the partitioning distribution (Fig. 8) does not
depend on \( c_t \), the compounds with solubilities below \( c_t \) will contribute relatively much more the
fitted \( \varepsilon_{ff} \) than for the previous case, and thus the more material there is in the “tail” of the
partitioning distribution, the worse the assumption about a single \( \varepsilon \) for the whole distribution.

To relate the theoretical work conducted here to realistic atmospheric organic aerosol
mixtures, Fig. 13 displays an example of a solubility distribution representing SOA formed
from dark ozonolysis of \( \alpha \)-pinene (Chen et al., 2008). The solubilities have been estimated with
SPARC (see Wania et al., 2014 and references therein). The average molar masses and O:C
ratios in each solubility bin are also displayed, along with the \( c_t \) values corresponding to the
activation points with limited solubility—assuming that the organics form an ideal mixture with
each other. Most of the material is predicted to have solubilities between 1 and 100 g L\(^{-1}\),
indicating that this fresh \( \alpha \)-pinene SOA is at the critical range of solubilities for limited
dissolution at the point of activation. This in turn suggests that the observed difference between
the \( \kappa \) values inferred from hygroscopicity and CCN-activation for this mixture might largely
result simply from the distribution of solubilities present.

4 Discussion and conclusions

We have studied the relationship between CCN activation and solubility of 144 different
theoretically constructed complex organic mixtures using Köhler theory, accounting for the
partial solubility of the compounds in water and assuming ideal interactions between the
dissolved molecules and water. The mixtures encompassed a wide variety of solubilities, and
were represented by solubility distributions with various solubility ranges and shapes
(analogously to the volatility basis set, VBS). Two limiting assumptions (ideal mixture vs. unity
activity) about the interactions between the organics in the insoluble organic phase were tested.
The results using this comprehensive solubility representation (termed as “the full model”) were
compared to commonly-used simplified descriptions of solubility: 1) assuming complete
dissolution; 2) representing the mixture with single hygroscopicity parameter \( \kappa \); 3) representing
the mixture with a single soluble fraction $\varepsilon_{\text{eff}}$. The calculations were carried out for particle dry sizes ranging from 20 to 500 nm and supersaturations between 0.03% and 5%, thus probing an atmospherically representative parameter space and resulting in total 5957 unique activation points.

Comparing the full model predictions to the simplified solubility descriptions, we find that assuming complete dissolution under-predicts the activation diameter up to about a factor of two for the studied mixtures. Our results indicate that about 70-80% of the material needs be dissolved at the point of activation for the complete dissolution assumption to predict activation diameters that are within 10% of that produced by the full solubility treatment. Adding a single parameter to describe the mixture solubility improved the situation considerably: the predictions of activation diameters based on a single $\varepsilon$ or $\kappa$ for a given mixture were within 10% of the full model predictions, the difference between these two approaches being only marginal.

The fitted soluble fractions, $\varepsilon_{\text{eff}}$, describing the solubility distribution (and thus the fitted $\kappa$ which is directly proportional to $\varepsilon$) were found to correspond well to the fraction of dry particle material with solubilities larger than a given threshold solubility $c_t$. For the ideal organic mixture assumption the median $c_t$ was 10 g L$^{-1}$, most of the values falling between 1 and 100 g L$^{-1}$, depending somewhat on the supersaturation. Since the material with solubilities outside this range can generally be treated as completely soluble or insoluble in CCN activation calculations, the error made by using the single soluble fraction increased when a larger fraction of material was present in this critical range. For the unity activity case the median $c_t$ was 1 g L$^{-1}$, but decreased with the number of components present in the mixture, $n$. For the range of $n = 3$-$100$ studied here, the typical $c_t$ values were between 0.1 and 10 g L$^{-1}$. Due to the asymmetric shape of the aqueous-organic phase partitioning of the organics in the unity activity case, the simplified models performed better, the more material with solubilities larger than $c_t$ was present in the particles. In general, the median values for $c_t$ represented the soluble fraction with a reasonable accuracy in most of the studied mixtures, although the exact composition of the mixtures varied considerably.

Our values for the limiting solubilities for complete dissolution are in agreement with the values of 3 g L$^{-1}$ and 1 g L$^{-1}$ previously reported by Huff Hartz et al. (2006) and Chan et al. (2008) based on experimental data on specific mixtures. Our results on the two different assumptions about the organic phase activities indicate that the mixtures investigated by these past studies were probably somewhat non-ideal, where the compounds hindered each other’s
dissolution less than would be expected for a fully ideal mixture. On the other hand, in light of our findings, the observations of the close to complete dissolution of SOA at activation (Huff Hartz et al., 2005; Engelhart et al., 2011) indicate that the majority of the material in the studied SOA mixtures had solubilities larger than 10 g L\(^{-1}\). Our results suggest that even with vastly different solubility distributions one can yield very similar CCN-activation behavior (and consequently values of \(\kappa\) or \(\varepsilon\)), as the parameter that matters is the material above \(c_t\).

The above results suggest that the solubility range corresponding to limited solubility in CCN activation is between 0.1-100 g L\(^{-1}\), and resolving the solubility distributions of aerosol mixtures outside this range provides little added value for understanding their CCN activation. In fact, this is probably a conservative estimate, as in most cases most material below 1 g L\(^{-1}\) is practically insoluble and most material above 10 g L\(^{-1}\) completely soluble – even considering the uncertainty in the organic mixture activity. These results can be used to guide the representation of the cloud activation properties of complex mixtures, and provide quantitative support for the previous notion that knowing the water-soluble fraction of the aerosol mixture in question is the key in most applications. We provide quantitative estimates on how this soluble fraction should be defined in the case of complex mixtures, and when such a simplified model is not expected to perform well.

There are, however, some limitations to our approach to keep in mind when applying the results to laboratory experiments or atmospheric data. Since the focus of this work was strictly on the links between solubility and CCN activation, we did not explore in depth how variation of surface activity, molecular mass, pure-component density, the gas-droplet partitioning of the organic compounds or non-ideality of water with respect to the aqueous phase would affect the results (see also Suda et al. 2012, 2014; Topping et al., 2012). Furthermore, temperature was assumed to stay constant at 298 K. Since many of the thermodynamic properties relevant to CCN activation are temperature-dependent (see e.g. Christensen et al., 2012), future work investigating the impact of temperature on the phenomena studied here is needed. Furthermore, the solubility and organic phase activity should naturally be linked to the aqueous phase activity coefficients predicted and used in a number of previous studies (see e.g. Topping et al., 2013 and references therein), although lack of well-defined experimental data on organic phase activities and mixture solubilities currently hinders quantitative evaluation of the current multi-phase mixture thermodynamic models (see e.g. Cappa et al., 2008). Evaluation of the concepts and approaches presented here (e.g. the
Another important future step would be applying the introduced solubility distribution framework in the atmospheric context. On one hand, the framework is likely to be useful in modeling the evolution of the CCN-activity of secondary organics. We expect the solubility distributions (and thus $c_t$) to depend on the SOA mixture properties such as the O:C ratio and the molar masses of the mixture constituents, which in turn evolve due to atmospheric chemistry (Kuwata et al., 2012; Suda et al., 2014), coupling the solubility distributions to the different dimensions of the VBS (Donahue et al., 2006; Donahue et al., 2011; Kroll et al., 2011; Donahue et al., 2012; Shiraiwa et al., 2014). A systematic study investigating the interlinkages between these variables in light of the available experimental data from field and laboratory would thus be a valuable future contribution. Furthermore, as atmospheric aerosol particles are typically mixtures of organic and inorganic constituents, the molecular interactions between atmospheric organics and inorganics as well as their effect on the pure-component solubility should be expanded. On the other hand, the solubility distribution can be used as a simplifying concept aiding in large-scale model simulations coupling atmospheric chemistry with the dynamics of cloud formation. With the assumptions applied here the CCN-activation calculation itself is computationally relatively light, slowing down considerably only if $n$ is of the order 10000 or larger with typical present-day computational resources. Therefore, using the solubility distribution framework within an atmospheric model is a good option if accuracy beyond the simple one-parameter approaches is required, or as an intermediate tool linking atmospheric age to the effective $\varepsilon$ or $\kappa$ describing a given mixture.

Acknowledgements

Financial support from the European Research Council project ATMOGAIN (grant 278277), Vetenskapsrådet (grant 2011-5120), the European Commission FP7 integrated project PEGASOS (grant 265148) are gratefully acknowledged.

References


atmospheric aerosol particles measured with H-TDMA instruments in various environments a
Topping, D. O., McFiggans, G. B.: Tight coupling of particle size, number and composition in
Topping, D.O., Barley, M., and McFiggans, G.: Including phase separation in a unified model
to calculate partitioning of vapours to mixed inorganic–organic aerosol particles, Faraday
Discuss., 165, 273, 2013.
VanReken, T. M., Ng, N. L., Flagan, R. C., and Seinfeld, J. H.: Cloud condensation nucleus
activation properties of biogenic secondary organic aerosol, J. Geophys. Res., 110, D07206,
Varutbangkul, V., Brechtel, F. J., Bahreini, R., Ng, N. L., Keywood, M. D., Kroll, J. H., Flagan,
aerosols formed by oxidation of cycloalkenes, monoterpenes, sesquiterpenes, and related
Volkamer, R., Jimenez, J. L., San Martini, F., Djepina, K., Zhang, Q., Salcedo, D., Molina, L.
air pollution: Rapid and higher than expected, Geophys. Res. Lett., 33, L17811,
gas-particle partitioning during the formation of secondary organic aerosol, Atmos. Chem.
Wex, H., Petters, M. D., Carrico, C. M., Hallbauer, E., Massling, A., McMeeking, G. R.,
hygroscopic growth and activation for secondary organic aerosol: Part I – evidence from
Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M.,
Petäjä, T., Clémer, K., van Roozendael, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T.,
Shaiganfar, R., Beirle, S., Apituley, a., Wilson, K. and Baltensperger, U.: Comparison of
ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR
Table 1. Simplified descriptions of organic mixture solubilities.

<table>
<thead>
<tr>
<th>Mixture model</th>
<th>Number of components</th>
<th>Solubility</th>
<th>Other input parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete dissolution</td>
<td>1</td>
<td>$c_{\text{sat}} \rightarrow \infty$</td>
<td>$M_{\text{org}}, \rho_{\text{org}}$</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1</td>
<td>$\kappa(c_{\text{sat}}, M_{\text{org}}, \rho_{\text{org}})$</td>
<td>-</td>
</tr>
<tr>
<td>Soluble fraction $\varepsilon_{\text{eff}}$</td>
<td>2</td>
<td>$c_{\text{sat},1} \rightarrow \infty$  $c_{\text{sat},2} = 0$</td>
<td>$M_{\text{org}}, \rho_{\text{org}}$</td>
</tr>
</tbody>
</table>

$^1c_{\text{sat}}$ the solubility (saturation concentration) in aqueous solution.
**Table 2.** Solubility distributions of the organic mixtures considered in this study.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Shape</th>
<th>Number of components</th>
<th>([c_{\text{sat,min}}, c_{\text{sat,max}}]) (^2) (g L(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flat, log (c)-axis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Flat, linear (c)-axis</td>
<td></td>
<td>Low: ([10^{-5}, 10^3])</td>
</tr>
<tr>
<td>3</td>
<td>Log. increasing</td>
<td>3, 5, 10, 100</td>
<td>Mid: ([0.1, 10^3])</td>
</tr>
<tr>
<td>4</td>
<td>Linear increasing</td>
<td></td>
<td>High: ([10, 10^3])</td>
</tr>
<tr>
<td>5</td>
<td>Log. decreasing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Linear decreasing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)For all solubility distributions two assumptions about the organic phase activity coefficients:

(1) ideal mixture and (2) unity activity (see text for details).

\(^2\)\(c_{\text{sat}}\) the solubility (pure component saturation concentration) in aqueous solution.
**Table 3.** Properties of water and organic compounds used in Köhler curve calculations (see Eq. 1).

<table>
<thead>
<tr>
<th>Property (unit)</th>
<th>Water (kg m$^{-3}$)</th>
<th>Organic $i$ (kg mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>1000</td>
<td>1500</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.073</td>
<td>-</td>
</tr>
<tr>
<td>$M$</td>
<td>0.018</td>
<td>0.18</td>
</tr>
</tbody>
</table>

1. These properties were chosen based on literature on the effective molar masses and densities determined for laboratory SOA (Engelhart et al., 2008; Asa-Awuku et al., 2010), and assumed to be same for every organic compound $i$. 
The dry particle is assumed to consist of \( n \) organic compounds, each denoted with a subscript \( i \). The wet particle is assumed to consist of a dry organic (insoluble) phase and an aqueous phase with water and dissolved organics. The aqueous phase is assumed to be in equilibrium with the ambient water vapour. \( Y \) refers to mole fractions in the organic phase, \( X \) to mole fractions in the aqueous phase, and \( m \) to the masses of the organic constituents and water. \( c_{i,eq} \) refers to the equilibrium concentration of each organic compound in the aqueous solution and \( p_{w,eq} \) to the equilibrium vapour pressure of water above the aqueous solution.

**Figure 1.** Schematic of the conceptual model used in the equilibrium composition calculations.
Figure 2. Examples of solubility distributions used in the calculations for saturation concentrations ranging from 0.1 to 1000 g L\(^{-1}\). (a) Linear and logarithmic flat distributions; (b) Linear and logarithmic increasing distributions; (c) Linear and logarithmic decreasing distributions. The numbers of the distributions refer to the numbering in Table 2 (see Sect. 2.1.4).
Figure 3. a) Examples of Köhler curves for the flat logarithmically spaced solubility distribution (Distribution 1 in Table 2) with $n = 5$ and the high solubility range (Table 2). The dots indicate the point of activation, black indicating incomplete dissolution ($\varepsilon < 1$) and red complete dissolution ($\varepsilon = 1$). b) The activation points determined from the model calculations (in total 7200 Köhler curves, see Table 2), corresponding to in total 5957 points in the activation dry diameter vs. supersaturation space. Also the dependence of the dissolved fraction at the point of activation is illustrated.
Figure 4. The dependence of the activation diameter for four different supersaturations ($s = 1\%, 0.6\%, 0.3\%$ and $0.1\%$, see also Fig. 3b) on the solubility range for the solubility distributions outlined in Table 2 for $n = 5$, and the two assumptions about the organic phase activity.
Figure 5. The activation diameter calculated using the solubility distributions (Table 2, referred to as the full model) and the simplified dissolution descriptions (Table 1) a) complete dissolution assumption; b) the hygroscopicity parameter $\kappa$; c) the soluble fraction $\varepsilon_{\text{eff}}$ to describe the solubility of the organic mixture. The symbols correspond to the best fits to the full model data. The black line shows the 1:1 correspondence between the two data sets.
Figure 6. The performance of the simplified solubility representations (see Table 1) in predicting the activation diameter for a given supersaturation as compared with the full model. The black bars depict the 25- and 75-percentiles and the gray bars the 10- and 90-percentiles.
Figure 7. a) The fitted soluble fraction $\varepsilon_{\text{eff}}$ as function of the true dissolved fraction $\varepsilon$ for each considered mixture (see Table 2). Symbols: mean $\varepsilon$ over all activation points. Grey lines: the range of $\varepsilon$ values at the different activation points for a given mixture; b) The fitted $\kappa$ values as a function as function of the true dissolved fraction $\varepsilon$ for each considered mixture. The red dashed line denotes the limit of $\kappa_{\text{max}} = 0.15$ which applies to all the studied mixtures. 1:1 lines are also indicated.
Figure 8. The dissolution behavior of the organic mixture corresponding to Distribution 1 with \( n = 100 \) and the low solubility range (see Table 2) at the activation point for \( s_c = 0.1\% \) (see Fig. 4). The figures depict the distribution of material in each solubility bin between the aqueous and the insoluble organic phases for the two different assumptions about the organic phase activity. \( c_t \) refers to the 50%-point of the partitioning (Eqs. 24-25). a) The ideal organic mixture; b) the unity activity assumption.
Figure 9. The distributions of the $c_t$ values (i.e. the 50% partitioning point, see Eqs. 24-25) at the point of activation for all the considered mixtures (Table 2) and activation points, and the two assumptions about the organic phase activity. a) The ideal organic mixture; b) the unity activity assumption.
Figure 10. Sensitivity of the $c_t$ values to supersaturation, molar mass and number of components for Distribution 1 with the mid solubility range (see Table 2). Only points with limited solubility ($0 < \varepsilon < 1$) at the point of activation are included.
Figure 11. The relationship between the fitted dissolved fraction $\varepsilon_{\text{eff}}$ at the point of activation and the mass fraction over the median $c_t$ for all the considered mixtures and $s = 1\%$, 0.6\%, 0.3\% and 0.1\%. Closed symbols: the ideal organic mixture assumption (median $c_t = 10$ g L$^{-1}$). Open symbols: the unity activity assumption for the organics (median $c_t = 1$ g L$^{-1}$).
Figure 12. The solubility distribution properties best explaining the performance of the three simplified solubility models, illustrated with the norm of the relative deviation of $D_{\text{p,act}}$ as compared with the full model predictions for $s = 1\%$, 0.6\%, 0.3\% and 0.1\%. The performance of the complete dissolution assumption as a function of the mass fraction with solubilities below the median $c_t$ for a) the ideal organic mixture (median $c_t = 10 \text{ g L}^{-1}$) and b) the unity organic activity (median $c_t = 1 \text{ g L}^{-1}$) assumptions. The performance of the $\kappa$ model as a function of the mass fraction with solubilities c) between 1-100 g L$^{-1}$ for the ideal organic mixture assumption; d) below 1 g L$^{-1}$ for the unity organic activity assumption. The performance of the $\epsilon$ model as a function of the mass fraction with solubilities e) between 1-100 g L$^{-1}$ for the ideal organic mixture assumption; f) below 1 g L$^{-1}$ for the unity organic activity assumption. The points close to complete dissolution ($\epsilon \geq 0.8$) are shown with lighter grey than the rest of the points. The error bars represent the variability with supersaturation and particle size.
Figure 13. A) An example of a solubility distribution for SOA generated in dark-ozonolysis of α-pinene SOA. The expected composition has been taken from Chen et al. (2008), and the pure-component solubilities have been calculated with the SPARC prediction system (e.g., Wania et al., 2014 and references therein). “Misc.” refers to completely miscible components. B) The dependence of the $c_t$ value at the point of activation on supersaturation assuming an ideal organic mixture, with only points corresponding to limited solubility ($0 < \varepsilon < 1$) displayed.