

We thank the referees for their thoughtful and constructive comments which we believe have helped improve the manuscript substantially. We have revised the manuscript following the referees' suggestions. You can find answers to the referee comments (*in italics*) below with additions to the manuscript and supplement text (**in bold**). The added references are listed at the end. The line and figure numbers refer to the original manuscript.

### Anonymous Referee #1

This Technical Note by Hyttinen et al. introduces a new approach for the simulation of activity coefficients and aqueous solubilities of mono- and dicarboxylic acids using the COSMOtherm software. The COSMO-RS-DARE model extension is used in this work, which includes hydration and dimerization of the organic acids in water. The paper is rather technical and focused on the choices made in this application of COSMO-RS-DARE. This is appropriate for a technical note. However, I found it difficult to judge the added value of this work in terms of the broader application of such methods in an entirely predictive way, e.g. to generate activity coefficient data sets for more complex mixtures of organic acids that have not been studied experimentally. The manuscript does not offer much advice in this respect nor a discussion about the applications of the model. Adding a discussion about strengths and weaknesses of the introduced method would improve the paper. The writing and structure of this technical note are overall appropriate and the topic is of interest for the ACP community. However, as indicated by several of my general and specific comments below, there are a number of mistakes, inconsistencies and unclear statements that need to be addressed before publication can be recommended.

*Author's response: We have added discussion of the strengths and weaknesses, as well as suggestions for the applicability of COSMO-RS-DARE and COSMO-RS, as described in more detail below. We also changed the title from "Technical note: Estimating aqueous solubilities and activity coefficients of mono- and  $\alpha,\omega$ -dicarboxylic acids using COSMO-RS-DARE" to "Technical note: Estimating aqueous solubilities and activity coefficients of mono- and  $\alpha,\omega$ -dicarboxylic acids using COSMO-RS" in order to not emphasize the COSMO-RS-DARE method too much because majority of the results shown here are calculated using COSMO-RS.*

**Changes in manuscript (section 1, line 55): Most atmospherically relevant multifunctional compounds are not readily available for experimental determination of thermodynamic properties. Accurate theoretical estimates are therefore essential for advancing current aerosol process modeling to include more complex compounds and mixtures. Here, we demonstrate the applicability of COSMO-RS theory in calculating condensed-phase properties of atmospherically relevant organic compounds. Carboxylic acids are among the most abundant and well characterized organic compounds in the troposphere and are therefore a good compound class to use to validate the use of COSMO-RS in atmospheric research.**

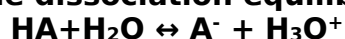
**(section 4, line 330): We showed that COSMOtherm provides a good solution to estimating thermodynamic properties of atmospherically relevant organic compounds that are not commercially available for measurements. In addition to simple binary systems studied here, COSMOtherm can be used to predict liquid-phase properties, such as activity coefficients, in complex, atmospherically relevant systems.**

## 1 General comments

• Aerosol acidity is mentioned in the introduction, but beyond that I could not find any description of the approach in COSMO-RS-DARE or COSMO-RS to account for partial dissociation of organic acids in solution. It would seem to be important as an effect that may compete with dimerization and hydration of undissociated acids. Please discuss.

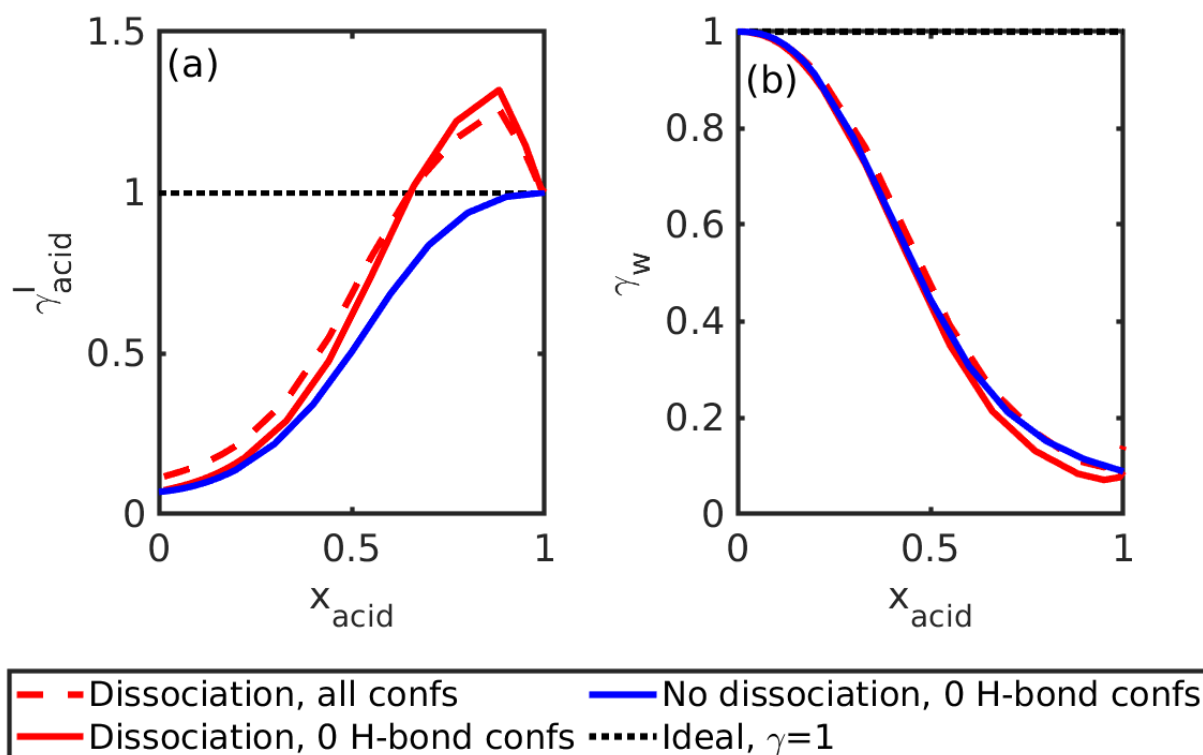
*Author's response: We agree with the reviewer that acid dissociation could be an important effect and its importance should be better described in the manuscript. A section about accounting for dissociation in calculation of oxalic acid and water activity coefficients was added.*

**Changes in manuscript (end of section 3.2.2, line 269): We additionally calculated activity coefficients of oxalic acid (the most acidic dicarboxylic acid of this study) with dissociation of oxalic acid included in the system. In this case, the system contains neutral oxalic acid (HA) and water (H<sub>2</sub>O), as well as deprotonated oxalic acid (A<sup>-</sup>) and hydronium ion (H<sub>3</sub>O<sup>+</sup>) according to the dissociation equilibrium**



**Figure S6 of the Supplement shows the difference between activity coefficients in a system where dissociation of oxalic acid is included and the binary system containing only neutral compounds. The calculation procedure is explained in more detail in the Supplement. There is no large difference in water activity coefficients when the ions are added to the system. A small change is seen in the acid activity coefficients, especially in the concentrated solutions where the estimated mole fraction of dissociated acid and hydronium ion is high. For the other carboxylic acids studied here, the effect of including dissociation is likely to be smaller than for oxalic acid, due to the lower mole fractions of ions present in solutions of less acidic compounds.**

**Changes in Supplement:**



**Figure S6: Activity coefficients of (a) oxalic acid and (b) water at 298.15 K calculated for systems where dissociation of oxalic acid was (red) and was not (blue) accounted for. On the x-axis, the mole fraction is for the total oxalic acid, meaning  $x_{\text{acid}}^{\text{DC}}$  in the systems where dissociation is included and  $x_{\text{acid}}$  in the system where dissociation is not taken into account. In both systems, the reference state for oxalic acid is the pure neutral acid and for water pure neutral water.**

• Fitting of reaction enthalpy parameters to existing experimental data was carried out, making the method perhaps less predictive than one would hope for. It is unclear how important the fitting of COSMO-RS-DARE model parameters is to achieve the presented activity coefficient and water-solubility results. If such fit parameters are essential, could you discuss the advantages of the COSMOtherm modelling approach compared to more traditional fitting of activity coefficient models for binary solutions, such as a Van Laar model or group-contribution models like UNIFAC?

*Author's response: We appreciate the comment and agree with the reviewer that the advantages of COSMOtherm modeling compared to group contribution methods should be included. We have added UNIFAC model estimates to the comparison between COSMOtherm-estimated and experimental activity coefficients.*

**Changes in manuscript (section 1, line 35): Group contribution methods, such as UNIFAC (Fredenslund et al., 1975) and AIOMFAC (Zuend et al., 2008), are often used to estimate activity coefficients of atmospherically relevant compounds.**

**(section 3.2.1, line 210): We additionally calculated UNIFAC predictions of acid and water activity coefficients using AIOMFAC-web (AIOMFAC-web, 2020; Zuend et al., 2008, 2011). These calculations (without inorganic ions) correspond to modified UNIFAC calculations by Peng et al. (2001). From Fig. S1 we see that, for acetic acid, the UNIFAC model underestimates the experimental activity coefficients more than even the COSMO-RS estimate.**

Similar to COSMO-RS, UNIFAC is not able to predict the increasing trend of water activity coefficients with the increasing acid mole fraction. (section 3.2.2, line 232): The water activity coefficients estimated using COSMO-RS are close to ones estimated using the UNIFAC model (modified UNIFAC; Peng et al. 2001). Similarly to what has been seen with the UNIFAC model, COSMOtherm is able to predict water activity coefficients at low acid mole fractions, while at high acid mole fractions both models fail to reproduce experimental activity coefficients. This indicates that COSMOtherm is not able to describe the water-acid interactions in supersaturated (crystalline) carboxylic acid mixtures.

(section 3.2.2, line 265): Comparing COSMOtherm (solid lines) and UNIFAC estimates (dotted lines), there is less variation between the UNIFAC-estimated activity coefficients for the different acids studied than between the COSMOtherm estimates. This indicates that, in COSMOtherm, the number of carbon atoms has a larger effect on activity coefficients than estimated by UNIFAC.

- The provided quantitative comparison to available measurements of dicarboxylic acid activity coefficients or water activities is rather limited and many existing data for concentrated aqueous solutions, e.g., by Choi and Chan, J. Phys. Chem. A 2002, 106, 4566-4572 or Marsh et al. (2017), could be used for a direct comparison of measured and predicted water activities / activity coefficients of the studied binary solutions. In addition, a comparison to predictions from other models, such as the UNIFAC / AIOMFAC / E-AIM models or other such approaches would allow the reader to compare the performance of COSMO-RS-DARE to such parameterized thermodynamic models that are often used in this community.

*Author's response: The missing experimental activity coefficients were added, in addition to UNIFAC and modified UNIFAC model estimations (Peng et al. 2001). Those are shown in Figs 5, 6, S2, S3 and S4. In addition, more discussion about the comparison between the different models and experiments was added.*

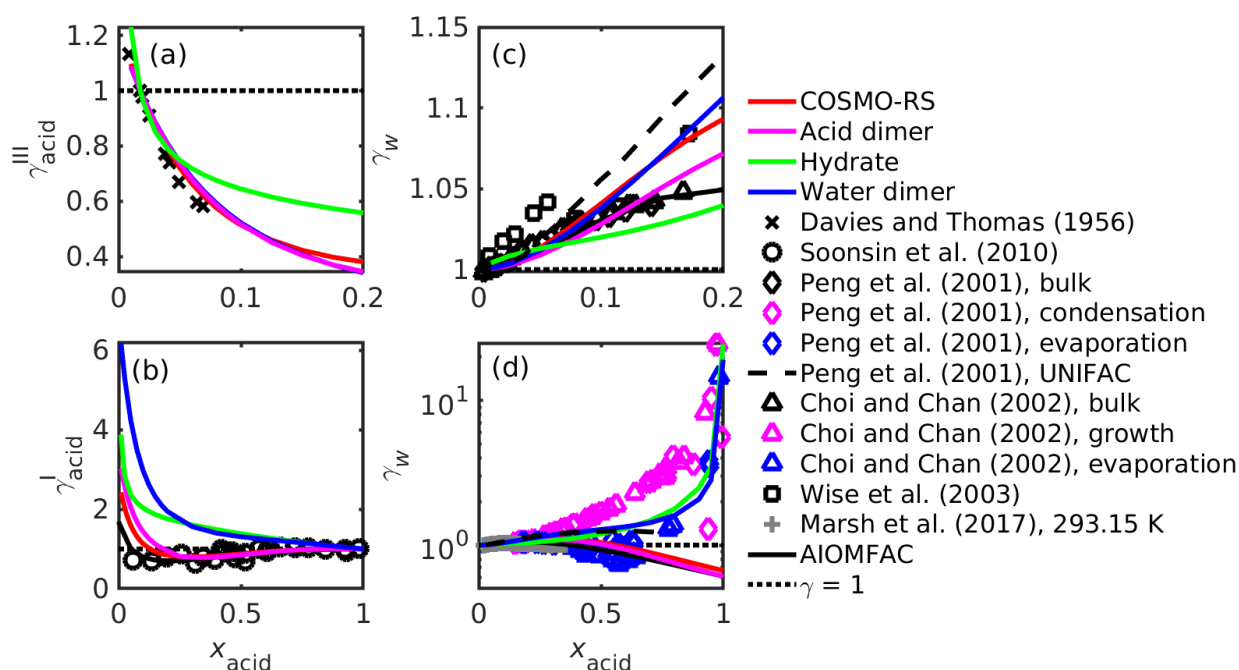
**Changes in manuscript (section 1, line 32): Activity coefficients of malonic, succinic and glutaric acid ( $m = 3, 4$  and  $5$ ) have been measured by Davies and Thomas (1956) and Soonsin et al. (2010) in bulk and particle experiments, respectively.**

**(beginning of section 3.2.2, line 232): We tested the effect of including different clusters in the activity coefficient calculation of malonic acid ( $m = 3$ ). A comparison between the experimental, UNIFAC-modeled and COSMOtherm-estimated activity coefficients are shown in Fig. 5. The malonic acid activity coefficients are compared in convention III (Fig. 5a) and in convention I (Fig. 5b). The COSMOtherm-estimated water activity coefficients are compared with experimental bulk (Fig. 5c) and particle (Fig. 5d) phase activity coefficients and UNIFAC-estimated activity coefficients. For malonic acid (and other studied dicarboxylic acids, see Figs S3-S5 of the Supplement), COSMO-RS-DARE is not able to improve the agreement between experiments and COSMOtherm estimates, the best overall fit is found using COSMO-RS.**

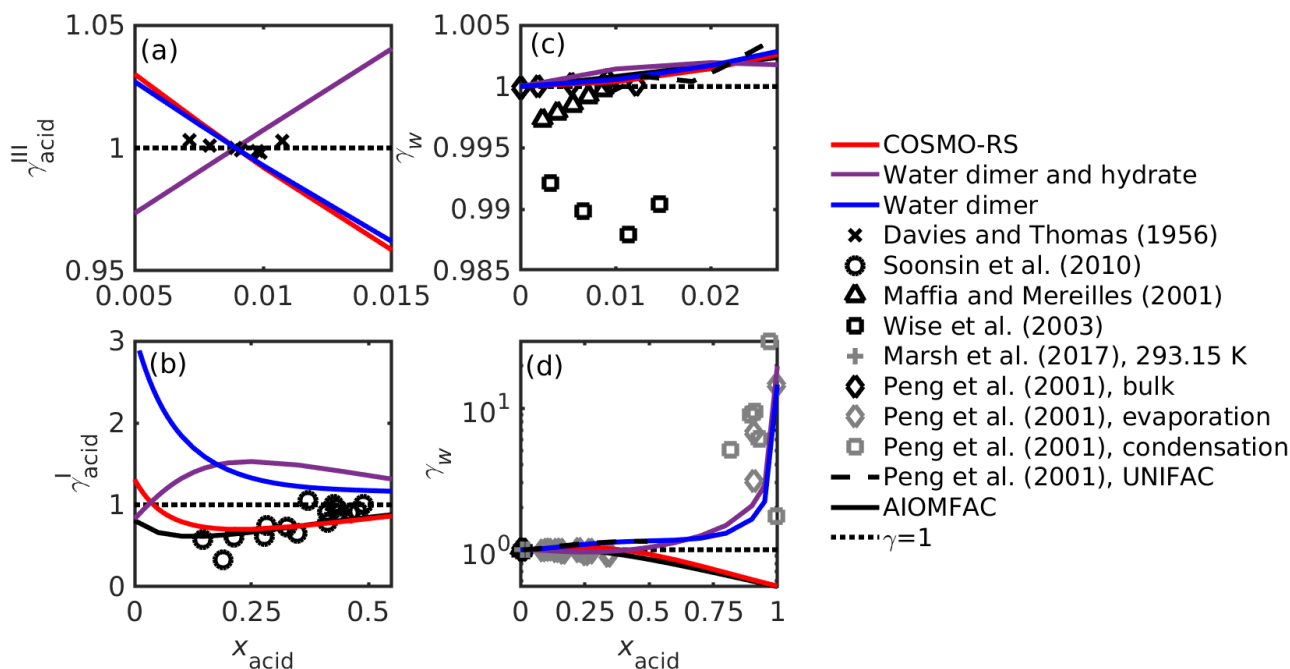
**(section 3.2.2, line 232): Figure S4 of the Supplement shows comparisons between experimental and COSMOtherm-estimated water activity coefficients in oxalic, adipic and pimelic acid. For these three acids, only water activities have been determined experimentally (Braban et al., 2003; Maffia and Mereilles, 2001, Marsh et al., (2017); Peng et al., 2001). In addition, water activities in adipic and pimelic acid solutions were only**

measured in bulk solutions (Marsh et al., 2017). We found a good agreement between the bulk measurements and COSMO-RS-estimated water activity coefficients, with COSMOtherm slightly overestimating the experiments. This result is in line with previous comparisons of hydroxy carboxylic acids (Hyttinen and Prisle, 2020).

### Changes in Supplement:



**Figure S3: Activity coefficients of (a-b) glutaric acid and (c-d) water in the binary mixtures at 298.15 K calculated using different clustering reactions in COSMOtherm calculation. As a comparison are experimentally determined activity coefficients of malonic acid by Davies and Thomas<sup>54</sup> (at 298.15 K given in convention III) and Soonsin et al.<sup>55</sup> (particle measurements at various temperatures given in convention I) and of water by Peng et al.,<sup>56</sup> Choi and Chan,<sup>57</sup> Marsh et al.,<sup>58</sup> Wise et al.<sup>59</sup> and AIOMFAC-web.<sup>53</sup>**



**Figure S4: Activity coefficients of (a-b) succinic acid and (c-d) water in binary mixtures at 298.15 K calculated using different clustering reactions in COSMOtherm calculations. As a comparison are experimentally determined activity coefficients of malonic acid by Davies and Thomas<sup>54</sup> (at 298.15 K given in convention III) and Soonsin et al.<sup>55</sup> (particle measurements at various temperatures given in convention I) and of water by Peng et al.,<sup>56</sup> Marsh et al.,<sup>58</sup> Wise et al.,<sup>59</sup> Maffia and Mereilles<sup>510</sup> and AIOMFAC-web.<sup>53</sup>**

### 1.1 Specific comments

• Abstract: The abstract would benefit from a quantitative statement about the average accuracy of the COSMO-RS-DARE predictions of activities or activity coefficients compared to the available experimental data and/or predictions by the standard COSMOtherm / COSMO-RS model. It would also be useful to state whether the outlined COSMO-RS-DARE method is fully predictive or not.

*Author's response: We agree that it would be beneficial to explicitly state whether the COSMO-RS-DARE method is fully predictive. The following has been added to the abstract:*

**Changes in manuscript (abstract, line 3): Conductor-like Screening Model for Real Solvents (COSMO-RS) underestimates experimental monocarboxylic acid activity coefficients by less than a factor of 2 but experimental water activity coefficients are underestimated more especially at high acid mole fractions.**

**(abstract, line 5): COSMO-RS-DARE is not fully predictive, but fitting parameters found here can be used to estimate thermodynamic properties of monocarboxylic acids in other aqueous solvents, such as salt solutions. For the dicarboxylic acids, COSMO-RS is sufficient for predicting aqueous solubility and activity coefficients and no fitting to experimental values is needed. This is highly beneficial for applications to atmospheric systems, as this data is typically not available for a wide range of mixing states realized in the atmosphere, either due to feasibility of the experiments or to sample availability.**

- Line 31: Clarify the sentence with "the Gibbs-Duhem equation was fitted"; what exactly was fitted? To my knowledge, the fundamental Gibbs-Duhem equation has no dedicated fit parameters.

*Author's response: We apologize for the confusion here. The self-consistency of measured data was checked using the Gibbs-Duhem equation, while the experimental points were fitted to self-consistent functions. This sentence was changed to:*

**Changes in manuscript (section 1, line 31): In addition, Hansen et al. (1955) represented the experimental points using self-consistent activity coefficient functions.**

- L. 32 – 34: The sentence should be improved given that, among the cited references, both Peng et al. (2001) and Choi and Chan (2002) use electrodynamic balance measurements that cover mass fractions of the solute far beyond the dilute solution range of the dicarboxylic acids. Further, the COSMOtherm predictions provided in this work (Fig. 3) seem not to be compared to such experimental data sets, even though the authors are aware of them.

*Author's response: We agree that the comparison with experiments should be more transparent. Measurements for supersaturated aerosol solutions were added to the activity coefficient comparisons. A figure of malonic acid and water activity coefficients compared to different experimental and UNIFAC values was moved to the main text. Comparison between experiments and COSMOtherm calculations of other compounds are shown in the Supplement. In addition, the text of section 2.3.2 was revised (see the response to the comment above on page 4-6).*

- L. 35: Define abbreviations for COSMOtherm and COSMO-RS.

*Author's response: Definition of COSMO-RS was added to line 35. COSMOtherm is a name of a program, not an abbreviation.*

- L. 70: Clarification necessary; Eq. (1), (2) define activity coefficients using the pseudo-chemical potential, which differs from the "regular" chemical potential more often used. However, comparison of Eqs. (1) and (2) raises the question how  $\mu_i^{*,li}$  in Eq. (1) differs from  $\mu_i^{*,li}$  in Eq. (2)? Eq. (2) seems to express the same relationship as Eq. (1). How exactly do they differ and what convention is used for the activity coefficients in Eq. (2)? Also, on line 71, the gas constant is expressed using kcal for energy; use of SI units would generally be preferred.

*Author's response: Eq (2) is the definition of pseudo-chemical potential, where the reference state is not specified. If the same reference state is used,  $\mu_i^{*,o}$  and  $\mu_i^o$  are equal. We added the following to clarify the definition. All energy units were converted to kJ/mol.*

**Changes in manuscript (section 2.1, line 71): By definition, the activity coefficient of a compound at the reference state is unity ( $\gamma_i^l(x_i=1)=1$ ), which leads to equal chemical and pseudo-chemical potential at the reference state. At other states ( $x_i < 1$ ), the relation between chemical and pseudo-chemical potentials ( $\mu$  and  $\mu^*$ , respectively) can be expressed as**

$$\mu_i^*(x_i) = \mu_i(x_i) - RT \ln x_i$$

• L. 76 – 79: Clarify the basis for Eq. (3), why should that apply (references /reasons)? Also, Eq. (4) stated in the current form seems incorrect and awkward: why write the right hand side composition  $a_{i,\beta}(1-x_{SOL,w})$ ? This seems not entirely correct and to be a potential source of confusion. This confusion exists because  $(1-x_{SOL,w})$  should be exactly the same as  $x_{SOL,acid}$  in a binary mixture (in the same phase), yet the former expression would only be correct for binary aqueous mixtures, not in general. However, in the LLE case, the mole fractions of acid in phases  $\alpha$  and  $\beta$  will differ, which is missing in Eq. (4). Why not write  $a_{i,\beta}(x_{\beta,SOL,acid})$  and analogously for phase  $\alpha$ . Further, please define the meaning of subscript SOL.

*Author's response: The specific compounds were removed from the definition of LLE to make it more general and the derivation of Eq. (3) was added. Furthermore, the definition of  $x_{SOL}$  was added.*

**Changes in manuscript (section 2.2, line 74): In LLE, the standard chemical potential ( $\mu$ ) of a compound is equal in both of the liquid phases ( $\alpha$  and  $\beta$ ):**

$$\mu_i(x_i^\alpha) = \mu_i(x_i^\beta)$$

**The standard chemical potential of compound i in a solution is defined using the standard chemical potential at the reference state:**

$$\mu_i(x_i) = \mu_i^\circ(x^\circ, T, P) + RT \ln a_i(x_i),$$

where  $a_i(x_i)$  is the activity of compound i with mole fraction  $x_i$ . The liquid-liquid equilibrium condition between the solvent-rich phase ( $\alpha$ ) and the solute-rich phase ( $\beta$ ) becomes:

$$a_i(x_i^\alpha) = a_i(x_i^\beta)$$

**Changes in manuscript (section 2.2, line 82): ... where  $x_{SOL,i}$  is the mole fraction solubility (SOL) of compound i in the solvent.**

• L. 98: The following statement "by multiplying the reaction equilibrium constant with the ratio of the activity coefficients..." and need for Eq. (11) seem unwarranted and require further explanation. Why should Eq. (11) be necessary? Is this because the authors are only considering mole fractions in the reaction constant, not activities? It is unclear because based on Eq. (8) – (10), in which chemical potentials and therefore implied activity coefficients are used, there seem to be no need for Eq. (11). Are you instead using pseudo-chemical potentials? Please clarify this. Also, activity coefficients of the reactants and products in what mixture? If the mixture contains a solvent, e.g. water, which affects the values of activity coefficients, how would Eq. (11) become solvent-independent?

*Author's response: We agree that this should be formulated more clearly in the manuscript. The definition of effective equilibrium constant was rewritten without using the equilibrium constant.*

**Changes in manuscript (section 2.3, line 91): COSMOtherm estimates effective equilibrium constants of condensed-phase reactions from the free energy of the reaction ( $\Delta G_r^{lo}$ ):**

$$K_{eff} = \exp(\Delta G_r^{lo} / RT)$$



**The reaction free energy is calculated from the free energies of the pure reactants ( $G_{\text{react}}^{\text{lo}}$ ) and products ( $G_{\text{prod}}^{\text{lo}}$ ):**

$$\Delta G_r^{\text{lo}} = \sum G_{\text{prod}}^{\text{lo}} - \sum G_{\text{react}}^{\text{lo}}$$

**The free energy of compound i is the sum of the energy of the solvated compound ( $E_{\text{COSMO}}$ ), the averaged correction for the dielectric energy (dE; Klamt et al., 1998) and the pseudo-chemical potential of the pure compound:**

$$G_i^{\text{lo}} = E_{\text{COSMO},i} + dE_i + \mu_i^{*,\text{lo}}(x^{\circ}, T, P)$$

• L. 114: "equilibrium constants on a mole fraction basis" – Unclear what "mole fraction basis" should imply here. If the equilibrium constant is computed based on free energies of reaction or chemical potentials of reaction, they are always dimensionless – as any thermodynamic equilibrium constant should always be (concentration-product-based equilibrium constants are only approximations and not thermodynamically correct).

*Author's response: We agree with the reviewer, the phrase "mole fraction basis" was removed from the manuscript.*

• L. 130 – 133: Eq. (15) and text: It is unclear why enthalpic and entropic energy contributions should not already be accounted for by  $\Delta_0$ . Also why is there a factor 2? It seems possible that the energy difference is just not known/predicted well enough, such that a fit parameter was introduced to match experimental data. Is that the motivation for the "enthalpic" contribution in Eq. (15)? Please discuss.

*Author's response: These fitting parameters were used since the COSMOtherm program cannot determine the energy of the monomers in the clusters. The calculation of the interaction energy has been further clarified in the manuscript.*

**Changes in manuscript (section 2.4, line 128): The formation free energy of the cluster ( $G(\text{A},\text{A}\cdot\text{B})$ ) is calculated using fitting parameters  $c_H$  and  $c_S$  (enthalpic and entropic contributions, respectively) to describe the interaction between the monomers A and B in the cluster ( $\text{A}\cdot\text{B}$ ):**

$$G(\text{A},\text{A}\cdot\text{B}) = c_H - c_S T,$$

**The fitting parameters are used because COSMOtherm is unable to calculate the energy of a monomer in a cluster. Instead, the energy of a monomer in a cluster is assumed to be equal to the energy of the lowest-energy conformer of the same compound and the favorability of the cluster formation is estimated using the fitting parameters. Without temperature dependent experimental data, it is not possible to fit both fitting parameters. We therefore consider the enthalpic parameter  $c_H$  as the total formation free energy parameter at 298.15 K, setting the entropic parameter  $c_S$  to zero.**

• L. 180 – 184: The discussion of hydration and dimerization in aqueous solutions raises the question whether the dissociation of the carboxylic acids into dissolved ions was considered in the simulation? In dilute aqueous solution, dissociation of the acid and formation of hydronium ions would seem to be favorable over non-dissociative hydration. Please discuss. Acidity and pH are mentioned in the introduction, but nothing is said about acid dissociation within COSMO-RS-DARE simulations.

*Author's response: We have added a comparison between activity coefficients of dissociated and non-dissociated oxalic acid (lowest  $pK_a$ ) to the Supplement. For further details see the response to the first comment above (page 2-3).*

• L. 198 – 199: Statement: "We used these experimental activity coefficients to fit the enthalpic parameters ( $\chi_H$ ) for each of the acids in the COSMO-RS-DARE calculations." Given that such fit parameters were introduced, how predictive is the outlined COSMO-RS-DARE method for calculation of activity coefficients in aqueous solutions for compounds where the experimental equilibrium constants or activity coefficients are not known? Elaborating the discussion on this key point seems important for applications in atmospheric aerosol modeling. For example, if these parameters were all ignored (set to zero), how different would the predictions be for activity coefficients and solubilities in water?

*Author's response: The manuscript text was revised to emphasize that fitting parameters are only needed to improve the activity coefficient and solubility calculations of monocarboxylic acids and dicarboxylic acids are described well by COSMOtherm without the COSMO-RS-DARE extension.*

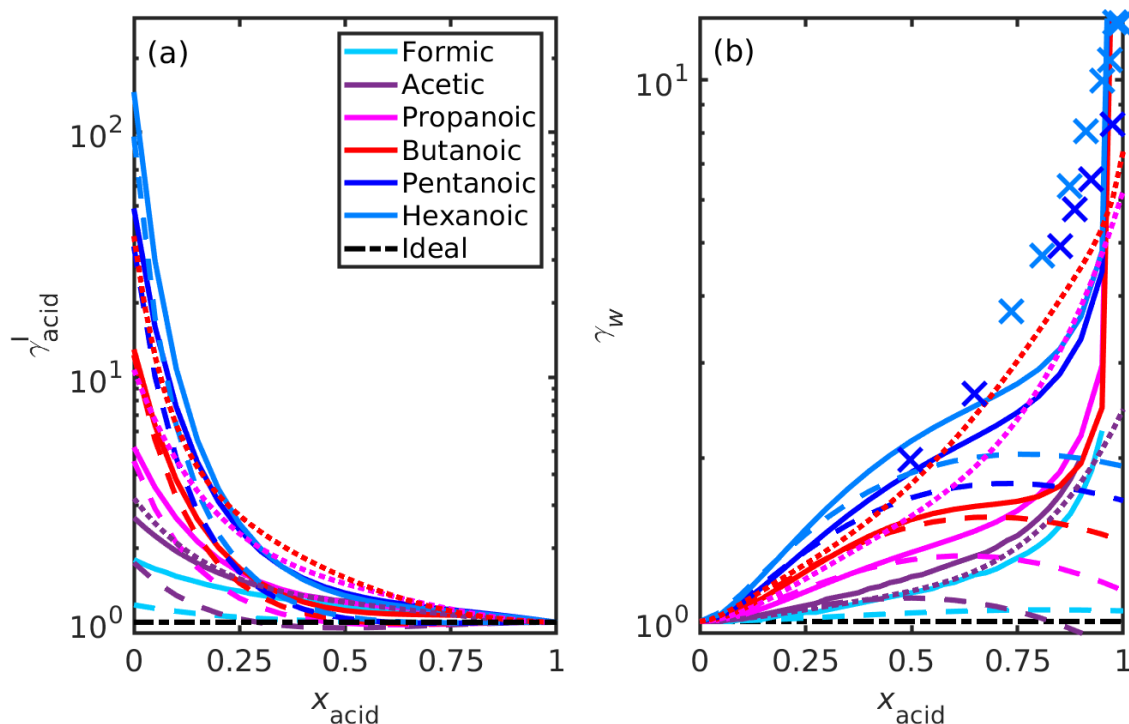
**Changes in manuscript (section 3.2.1, line 216): If the enthalpic parameters in COSMO-RS-DARE calculations are not fitted to experimental activity coefficients and instead are set to zero, the activity coefficients of both acid and water underestimate the experimental activity coefficients of Hansen et al. (1955) (see Fig. S2 of the Supplement).**

**(section 3.2.1, line 221): While COSMO-RS is fully predictive, COSMO-RS-DARE requires parameter fitting using experimental data. Fitted COSMO-RS-DARE parameters from one system can be used in other systems where the same clustering reactions are relevant. For instance, Sachsenhauser et al. (2014) found that the same interaction parameters of acid dimers can be used in systems containing other similar (non-polar) solvents. This indicates that our interaction enthalpies can be applied to other aqueous systems, for instance, ternary systems containing an inorganic salt, in addition to the carboxylic acid and water. This would allow for extending the findings of this study to atmospherically relevant aerosol solutions.**

**(section 3.2.1, line 216): If no experimental activity coefficients are available for fitting the COSMO-RS-DARE parameters, COSMO-RS estimates agree with experiments overall better than COSMO-RS-DARE or UNIFAC. COSMO-RS-estimated acid activity coefficients are close to the measured values in all mixing states, and for water activity coefficients the agreement between COSMO-RS and experiments is good in mixing states with  $x_{\text{acid}} < 0.75$ .**

**(section 4, line 307): We were also able to estimate activity coefficients of pentanoic and hexanoic acids using only experimental water activity coefficients in the fitting of the COSMO-RS-DARE enthalpic parameters.**

**Changes in Supplement:**



**Figure S2: Activity coefficients ( $\gamma^i$ ) of (a) monocarboxylic acids and (b) water in all mixing states of the binary aqueous solutions, at 298.15 K. The solid lines represent activity coefficient estimates using COSMO-RS-DARE ( $c_H=0$ ), dashed lines are UNIFAC estimates, dotted lines are calculated from the equations fitted to experiments by Hansen et al. (1955), and the markers are the experimental points from the same study.**

• L. 214, 215: Do these calculation include phase separation for the larger acids or is a mixed phase assumed for all compositions? At higher organic acid mole fractions, the binary mixtures show high values for water activity coefficients, as also stated in the caption to Fig. 2. This may suggest liquid-liquid phase separation could occur, which could affect the interpretation of the experimental data used for comparison. Please discuss.

*Author's response: Phase separation occurs in aqueous pentanoic and hexanoic acid mixtures based on experimental solubilities and COSMOtherm calculations. The experimental points by Hansen et al. (1955) shown in Fig. 5 for water activity coefficients in pentanoic and hexanoic acid were measured in mixing states corresponding to the acid-rich phases.*

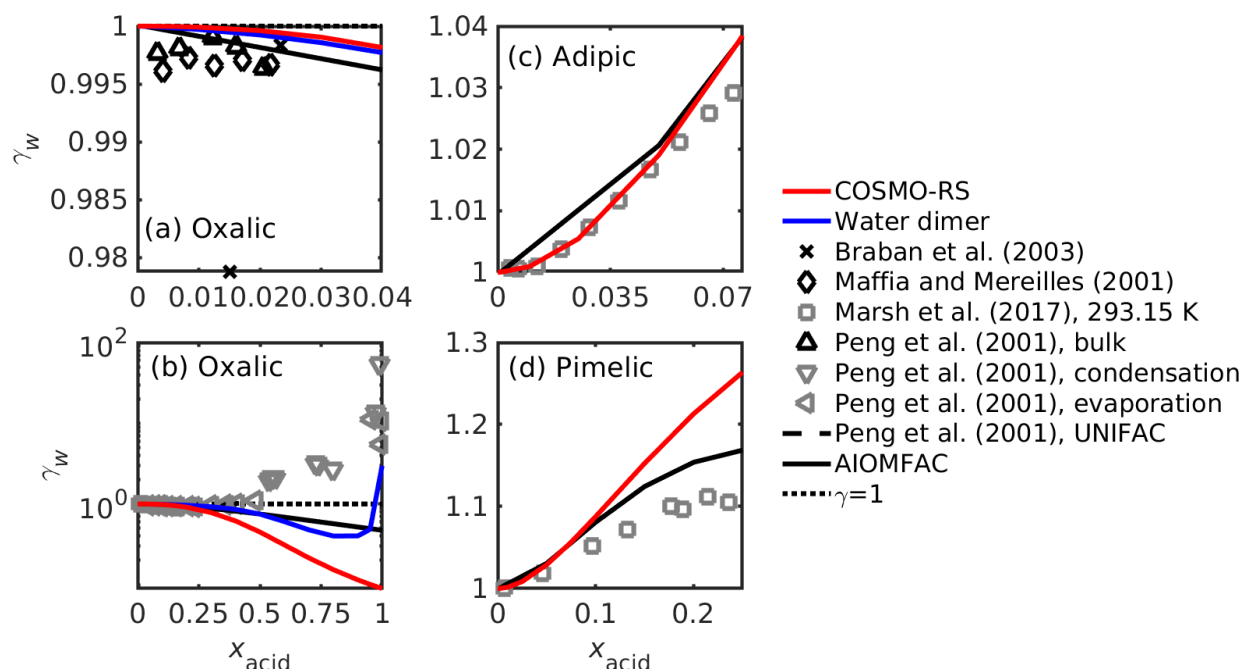
**Changes in manuscript (Figure 2 caption): For the studied acids with finite aqueous solubilities at 298.15 K (pentanoic and hexanoic acid), water activity coefficients were measured using acid-rich solutions (Hansen et al., 1955).**

• L. 261: "However, without experimental activity coefficient data of oxalic acid-water systems, we are not able to fit the  $c_H$  parameter needed for reliable estimates." – This statement is simply not correct. A literature search reveals that there exist multiple useful measurements for this binary system, such as water activities at dilute and concentrated conditions, from several sources. These include the work of Maffia and

Meirelles (2001) mentioned earlier in the study, as well as work by Braban et al. (2003) and Marsh et al. (2017), which provide water activity data that should be used here.

*Author's response: We highly appreciate the supplied references and a comparison of water activity coefficients in oxalic acid were added to the Supplement.*

### Changes in Supplement:



**Figure S5: COSMOtherm-estimated water activity coefficients in aqueous (a-b) oxalic acid, (c) adipic acid and (d) pimelic acid solutions at 298.15 K. The experimental and model activity coefficients are by Peng et al.,<sup>56</sup> Marsh et al.,<sup>58</sup> Maffia and Mereilles,<sup>510</sup> Braban et al.<sup>511</sup> and AIOMFAC-web.<sup>53</sup>**

• L. 264 and Fig. 3: The partial dissociation of dicarboxylic acids is either not considered at low acid concentrations or not discussed, even though dissociation would seem to be likely, especially for the smaller acid molecules. Was it determined to be irrelevant? This will require adequate discussion. Also, mention that mole fractions used in this work are defined assuming undissociated acid molecules (if this is indeed the case).

*Author's response: We agree that it is necessary to further clarify whether dissociation is taken into account. We added a comparison between activity coefficients in non-dissociated and dissociated oxalic acid-water systems, see response to the first general comment of referee #1 above (page 2-3). In addition, the following sentence was added to the manuscript:*

**Changes in manuscript (section 2.1, line 71): Unless otherwise mentioned, the mole fractions  $x_i$  correspond to mole fractions of undissociated acid or neutral non-protonated water.**

• L. 290 – 294: On melting temperatures and solids considered: "Cornils and Lappe (2000) and Omar and Ulrich (2006) also measured the melting point of oxalic acid and

found temperatures almost 100 K higher than Booth et al. (2010). ..." In this context, are the differences in melting points due to unclear statements about the crystalline form of oxalic acid (anhydrous or dihydrate)? Omar and Ulrich (2006) show in their paper that the solid-solid phase transition from the dihydrate to the anhydrous oxalic acid crystal polymorph occurs around 378 K, which is close to the melting temperature stated by Booth et al., while the anhydrous oxalic acid melts at 465 K. Therefore, the correct values of use in the COSMOtherm simulations will depend on the temperature range of interest; for room temperature, using the dihydrate form is very likely the crystalline solid to be considered and consequently the solubility equilibrium should be solved for oxalic acid dihydrate not the anhydrous form. It is also clear that the actual melting temperature and enthalpy of fusion is not as uncertain as the current text implies. Using the correct equilibrium relations for the dihydrate the agreement between COSMOtherm and the measured solubility data would be expected to be much better than shown in Fig. 3.

*Author's response: The melting point reported by Booth et al. (2010) was removed and the lower limit free energy of fusion was calculated using the melting point measured by Cornils and Lappe (2000) instead. In COSMOtherm, the SLE is an equilibrium between anhydrous solute and solvent. The transition considered for the free energy of fusion calculation is from anhydrous solid into anhydrous liquid.*

**Changes in manuscript (Table 1 caption): The melting point of oxalic acid reported by Booth et al. (2010) (370 K) is likely the temperature of the phase transition from dihydrate to anhydrous crystal polymorph. Similar transition was seen by Omar and Ulrich (2006) at 378.35 K in their differential scanning calorimetry experiment.**

• L. 305: "to significantly improve the activity coefficient estimates ..." - compared to what? COSMO-RS or other methods? What about a comparison to methods like UNIFAC, such as the model by Peng et al. (2001).

*Author's response: UNIFAC estimates by Peng et al. (2001) were added to the activity coefficient comparison figures (Figs 5, S1, S3-S5).*

**Changes in manuscript (section 4, line 305): We compared COSMOtherm-estimated activity coefficients and aqueous solubilities of simple carboxylic acids with experimental values and a commonly used UNIFAC model, and found a good agreement between experiments and COSMOtherm estimates. Using COSMO-RS-DARE, we were additionally able to improve the agreement between estimated and experimental water activity coefficient in binary monocarboxylic acid-water systems significantly compared to using COSMO-RS or UNIFAC.**

• L. 308: "In addition, COSMO-RS-DARE is able to predict the miscibility of butanoic acid in water, while COSMO-RS predicts a finite solubility" - This is nice. However, how much of that success comes from fitting model parameters ( $\chi_H$ ) rather than the use of the DARE extension?

*Author's response: The success of using COSMO-RS-DARE comes from both fitting the interaction parameters and selecting the appropriate clustering reactions for each calculation. Here, we have fitted the enthalpic parameter using activity coefficients of water and acid in the binary systems and used those parameters to estimate liquid-liquid equilibria. Fitting parameters are needed in the COSMO-RS-DARE method, otherwise the favorability of clusters formation is not modeled correctly. Similarly,*

COSMO-RS and group contribution methods are parametrized using experimental data. The difference between COSMO-RS-DARE and COSMO-RS is that COSMO-RS-DARE is not (yet) parametrized in COSMOtherm and the user needs to fit the parameters for each compound.

**Changes in manuscript (section 4, line 307): In addition, COSMO-RS-DARE was able to predict the miscibility of butanoic acid in water (using the fitting parameters of activity coefficient calculations), while COSMO-RS predicted a finite solubility.**

- Table 1: State whether the listed data are for the pure anhydrous solids of the acids or for hydrates (especially in case of oxalic acid). Also replace "literature values" by more appropriate wording.

*Author's response: We agree that the form of the solid should be clarified in the table. Furthermore, we have changed the formulation "literature values" to "experimental melting points".*

**Changes in manuscript (Table 1 caption): List of the studied  $\alpha,\omega$ -dicarboxylic acids and their experimentally determined melting points and heats of fusion. The values were measured using anhydrous acids (crystalline at 298.15 K).**

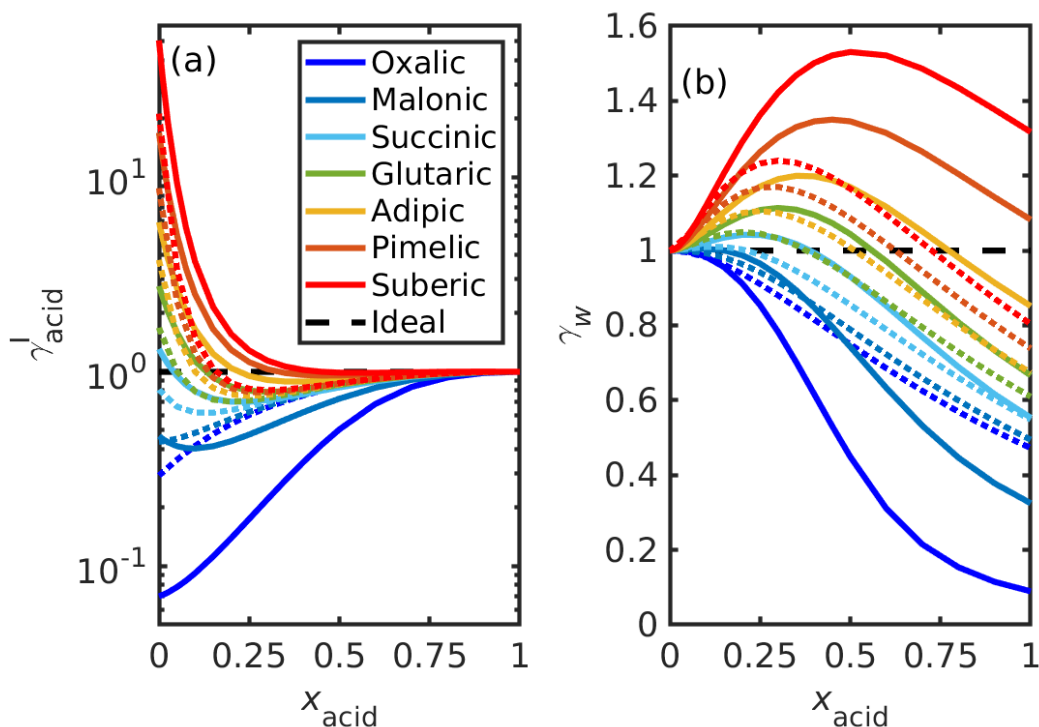
- Table 2: state the temperature for the listed equilibrium constants and for completeness also for which phase / solvent they apply (given that there is also gas-phase dimerization of such acids).

**Changes in manuscript (Table 2 caption): Dimensionless effective equilibrium constants ( $K_{\text{eff}}$ ) of cluster formation in condensed phase, at 298.15 K.**

- Figure 3: Why are the prediction data cut off at the solubility limits? It would seem useful to mark the solubility limit at 298.15 K for each acid, but to also show the predictions for the supersaturated range (which may apply in aerosols). This would also facilitate an extended comparison with experimental data existing for those higher concentrations, e.g. by Choi and Chan (2002) and Marsh et al. (2017). With the chosen log-scale for the x-axis, too much emphasis is put on the very dilute concentration range below  $10^{-3} x_{\text{acid}}$ , which seems not to be insightful.

*Author's response: To further improve clarity, the activity coefficients in Figure 3 were extended to the whole mole fraction range and the x-axis was changed to linear scale. Using linear scale on the x-axis, only solubilities of malonic and glutaric acid would be visible in the figure. We have therefore decided not to show the solubilities of the acids in this figure.*

**Updated figure in the main text (Figure 3):**



**Figure 6. COSMO-RS- (solid lines) and UNIFAC-estimated (dotted lines; AIOMFAC-web, 2020) activity coefficients ( $\gamma^i$ ) of (a) dicarboxylic acids and (b) water in the binary acid-water mixtures at 298.15 K. All COSMOtherm-estimated activity coefficient values are given in Tables S5 and S6 of the Supplement.**

## 2 References mentioned

- Braban, C. F., Carroll, M. F., Styler, S. A., and Abbatt, J. P. D.: Phase Transitions of Malonic and Oxalic Acid Aerosols, *The Journal of Physical Chemistry A*, 107,6594-6602, 10.1021/jp034483f, 2003.
- Choi, M. Y., and Chan, C. K.: Continuous Measurements of the Water Activities of Aqueous Droplets of Water-Soluble Organic Compounds, *The Journal of Physical Chemistry A*, 106, 4566-4572, 10.1021/jp013875o, 2002.
- Maffia, M. C., and Meirelles, A. J. A.: Water Activity and pH in Aqueous Poly-carboxylic Acid Systems, *Journal of Chemical & Engineering Data*, 46, 582-587, 10.1021/je0002890, 2001.
- Marsh, A., Miles, R. E. H., Rovelli, G., Cowling, A. G., Nandy, L., Dutcher, C. S., and Reid, J. P.: Influence of organic compound functionality on aerosol hygroscopicity: dicarboxylic acids, alkyl-substituents, sugars and amino acids, *Atmos. Chem. Phys.*, 17, 5583-5599, 10.5194/acp-17-5583-2017, 2017

## Anonymous Referee #2

### General comment

This paper employs the recently developed COSMO-RS-DARE model to estimate activity coefficients and solubilities of carboxylic acids in water. COSMO-RS-DARE is an extension of COSMOtherm that takes dimerization and aggregation in solution explicitly into account. This technical note concludes that COSMO-RS-DARE leads to better agreement with experimental data than COSMOtherm for the investigated mixtures. Although this paper is submitted as a technical note, the technical description of COSMOtherm and its extension COSMO-RS-DARE is lacking a proper

derivation and explanation. Also, the benefit of COSMO-RS-DARE compared with COSMOtherm remains unclear. As it seems, the new method relies on experimentally determined activity coefficients to calculate dimerization equilibria. Therefore, the benefit of COSMO-RS-DARE in the absence of experimental data is unclear. It is not clear whether COSMO-RS-DARE just performs better in predicting solubilities because of an additional degree of freedom introduced through potential dimerizations or a more accurate description of the system.

Major revisions of the manuscript are required before this technical note can be considered for publication. The different COSMO versions need to be explained better and the discussion of the results needs to be improved.

*Author response: Detailed descriptions and derivations of both COSMO-RS and COSMO-RS-DARE have already been published in papers cited in the work. In this paper, we test these methods for an atmospherically relevant chemical system, repeating the derivation here would be beyond the scope of the paper. However, we have improved the general explanation of the methods in the context of testing the methods, as described in more detail above as a response to comments from referee #1. In addition, we have clarified the comparison of COSMO-RS-DARE and COSMO-RS with other approaches and discussed the predictiveness of COSMO-RS-DARE, as described above as a response to referee #1. COSMO-RS-DARE was added to the title of section 2.4 to distinguish the COSMO-RS-DARE theory from COSMO-RS.*

#### **Changes in manuscript (section 2.4 title): Concentration dependent reactions (COSMO-RS-DARE)**

##### Specific comments

Lines 20 – 22: Here, acidity is mentioned as highly relevant. But the approach used in this technical note totally neglects deprotonation of acids.

*Author's response: Additional investigation of acid dissociation was added for the most acidic of the studied carboxylic acids, oxalic acid. See the response to the first general comment of referee #1 (page 2-3).*

Line 25: activity data of carboxylic acid-water systems is abundant as exemplified by the studies mentioned just below this sentence and there are even more. Please revise this sentence.

*Author's response: We apologize a slight mix-up here. This was meant to say "acid activity coefficients". This was added to the text.*

#### **Changes in manuscript (section 1, line 25): However, the acid activity data of carboxylic acid-water systems is much scarcer.**

Lines 65 – 66: the meaning of a pseudo-chemical potential should be explained.

*Author's response: More explanation for pseudo-chemical potential was added.*

#### **Changes in manuscript (section 2.1, line 71): Pseudo-chemical potential (Ben-Naim, 1987) is an auxiliary quantity defined using the standard chemical potential at the reference state $\mu^\circ$ :**

$$\mu_i^*(x_i) = \mu_i^\circ(x^\circ, T, P) + RT \ln \gamma_i(x_i)$$



**Pseudo-chemical potential has recently been used in molecular level solvation thermodynamics as a replacement to chemical potential (Sordo, 2015).**

Line 79: activity should be replaced by the activity coefficient in this equation.

*Author's response: The LLE condition used in the COSMOtherm calculations is that the activity of compound  $i$  is equal in both phases. The assumption is that there is no pure compound phase in the system. The equation was revised as a response to a comment from referee #1 (page 8).*

Line 81: This equation should be derived or a reference should be given.

*Author's response: The reference (Eckert and Klamt, 2019) was added to the equation.*

Line 96: How is the dielectric energy calculated or defined?

*Author's response: We have added a reference (Klamt et al. 1998) to the averaged correction to the dielectric energy.*

**Changes in manuscript (section 2.3, line 95): The free energy of compound  $i$  is the sum of the energy of the solvated compound (Ecosmo), the averaged correction for the dielectric energy (dE; Klamt et al., 1998) and the pseudo-chemical potential of the pure compound:**

Line 96: The difference between the chemical potential and the pseudo chemical potential is not clearly made and not explained. Here, the same symbol is used to refer to the chemical potential that was used before for the pseudo chemical potential.

*Author's response: We agree that it should be more transparent which chemical potential we refer to. The definition of pseudo-chemical potential as a function of chemical potential was added to the manuscript. In addition, "chemical potential" was replaced by "pseudo-chemical potential" in appropriate places to clarify that pseudo-chemical potential is the one that is used in all calculations.*

Line 100: Equation (11) needs to be explained better.

Line 112: The derivation of Eq. (14) remains obscure. The equation rather seems to be a definition of the effective equilibrium constant than a derived equation.

*Author's response: The definition of effective equilibrium was simplified. See response to a comment of referee #1 (page 8-9).*

Line 116: How is the surface of a molecule defined? Either explain here or give a reference.

**Changes in manuscript (section 2.4, line 117): The surface is considered as an interface between a virtual conductor around the molecule and the cavity formed by the molecule (Klamt and Schüürmann, 1993).**

Line 125: what is a property calculation?

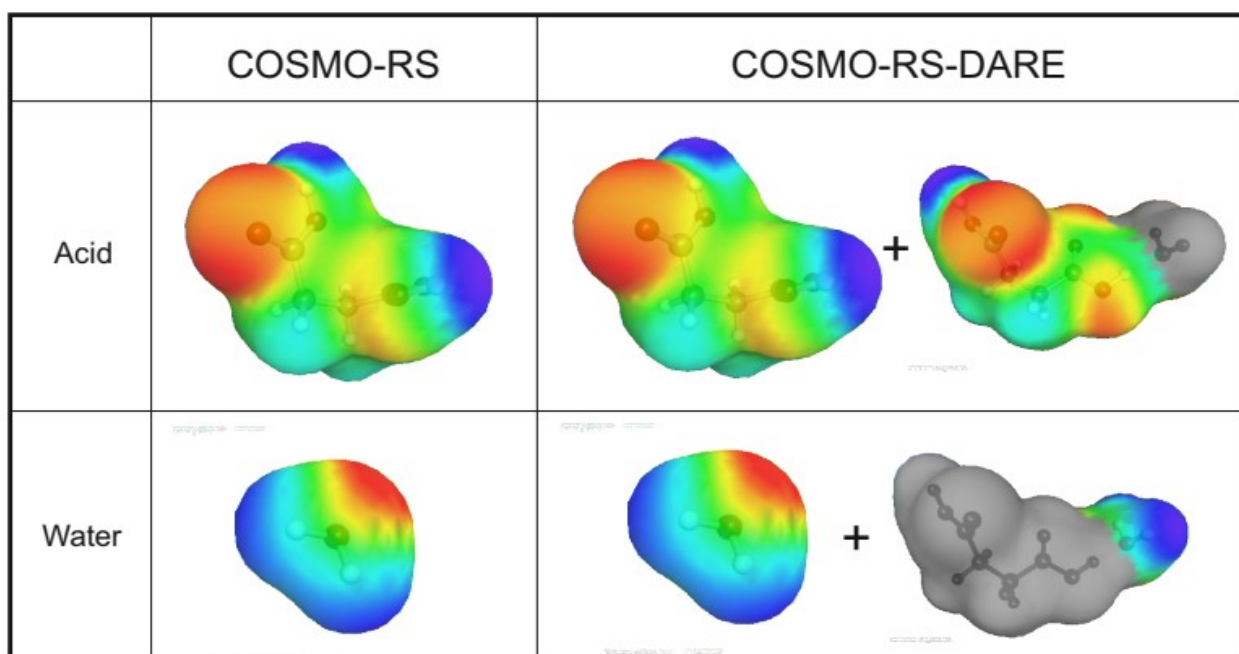
*Author's response: By "property calculation" we mean a calculation of any thermodynamic property in COSMOtherm.*

**Changes in manuscript (section 2.4, line 125): “property calculation” was changed to “COSMOtherm calculation”**

Line 125: what is the screening charge density? A scheme might help to explain it.

Author's response: We thank the referee for this suggestion, a scheme clarifying screening charge densities was added to the manuscript.

**Changes in manuscript (section 2.4, line 117): Each surface segment has an area ( $[\text{\AA}^2]$ ) and a screening charge density ( $\sigma[e \text{\AA}^{-2}]$ ).**



**Figure 1. The  $\sigma$ -surfaces of succinic acid and water conformers used in COSMO-RS and COSMO-RS-DARE calculations. The conformer distributions in COSMO-RS-DARE include parts of cluster  $\sigma$ -surfaces (in this example a hydrate cluster). Color coding of  $\sigma$ -surfaces: red = negative partial charge, blue = positive partial charge, green = neutral partial charge, grey = omitted  $\sigma$ -surface.**

Line 128 - 129: Why are interaction sites of molecule B not treated the same way?

Author's response: Molecule B is also included in the COSMO-RS-DARE calculations, the text was mistakenly left out of the manuscript. This has been corrected.

**Changes in manuscript (section 2.4, line 126): Similarly, the clustering product of molecule B is included in the calculation by omitting the  $\sigma$ -surface assigned to molecule A from the  $\sigma$ -surface of A·B. Examples of these partial  $\sigma$ -surfaces are shown on the right hand side of Fig. 1.**

Line 133 - 134: Why is the entropic parameter kept zero? This seems arbitrary. Please justify.

Author's response: We use only the enthalpic contribution to describe the total energy contribution, because we do not have temperature dependent measurements to fit both enthalpic and entropic parameters. The same result would be achieved if we

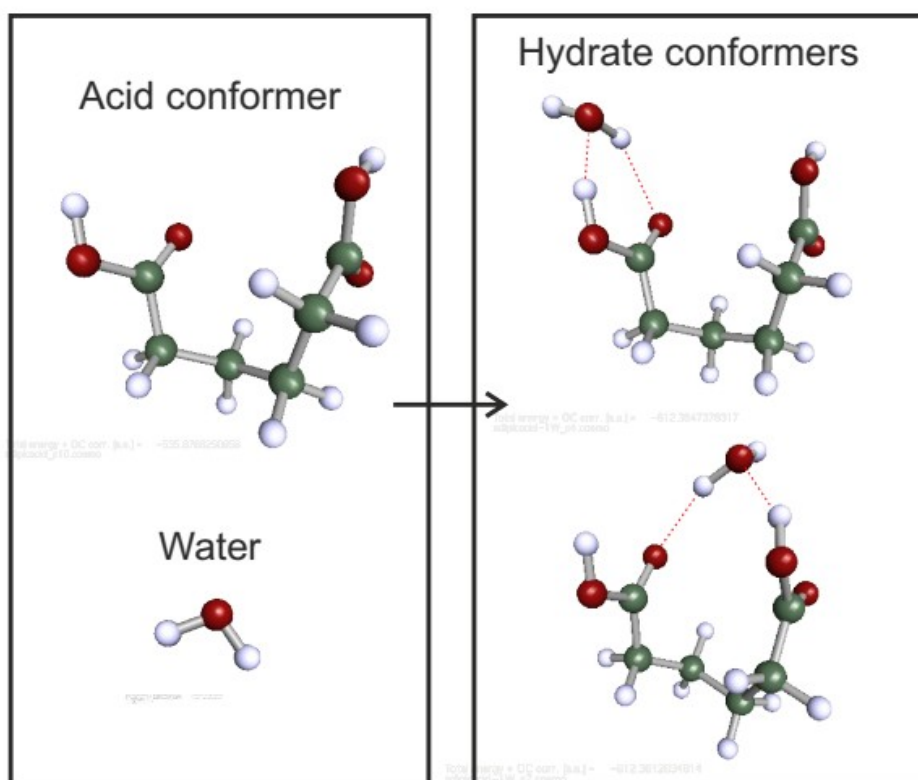
gave the entropic parameter any value so that  $c_H - c_S T$  equals the enthalpic parameter we found in the fitting. The enthalpic parameter is only valid in the temperature of our calculations (298.15 K), for other temperatures the entropic parameter must be fitted separately using temperature dependent measurements. This has been further clarified in the text.

**Changes in manuscript (section 2.4, line 131): Without temperature dependent experimental data, it is not possible to fit the entropic parameter. We therefore consider the enthalpic parameter  $c_H$  as the total formation free energy parameter at 298.15 K, setting the entropic parameter  $c_S$  to zero.**

Lines 150 - 158: This section is difficult to understand. A scheme might help.

Author's response: We agree that a scheme will aid the reader. The following figure has been added to the manuscript.

**Changes in manuscript:**



**Figure 2. The formation of dicarboxylic acid hydrate conformers. Color coding: green = C, white = H, red = O.**

Lines 225 - 227: This finding questions the benefit of the method.

Author's response: The method still has its benefits. When  $x_{water} \rightarrow 0$ , an important part of the description of water is removed from the calculation (water in hydrate), since the system contains no water and thus no hydrates or water dimers. The method can still be applied to estimate activity coefficients of water in solutions where  $x_{acid} < 0.9$ .

**Changes in manuscript: However, when the hydrate and water dimer reactions are included, COSMO-RS-DARE is not able to predict realistic**

activity coefficients for water at high mole fractions ( $x_{\text{acid}} > 0.9$ ) of the acids. This is likely due to the low concentration of water in the binary solution, leading to errors in the description of the interactions between water molecules. Still, COSMO-RS-DARE estimates agree well with the experiments at least up to 0.9 mole fraction of the monocarboxylic acids. This is an improvement compared to the UNIFAC model, which fails to reproduce experimental water activity coefficients already at acid mole fractions above 0.25. At very high acid mole fractions ( $x_{\text{acid}} > 0.95$ ), COSMO-RS-DARE predicts several orders of magnitude higher activity coefficients than what was seen in experiments.

Lines 234: I would not refer to dicarboxylic acids as being of low aqueous solubility. Some dicarboxylic acids have a high solubility. Moreover, data well into the supersaturated range is available (e.g. in Soonsin et al., 2010). This sentence needs to be revised accordingly.

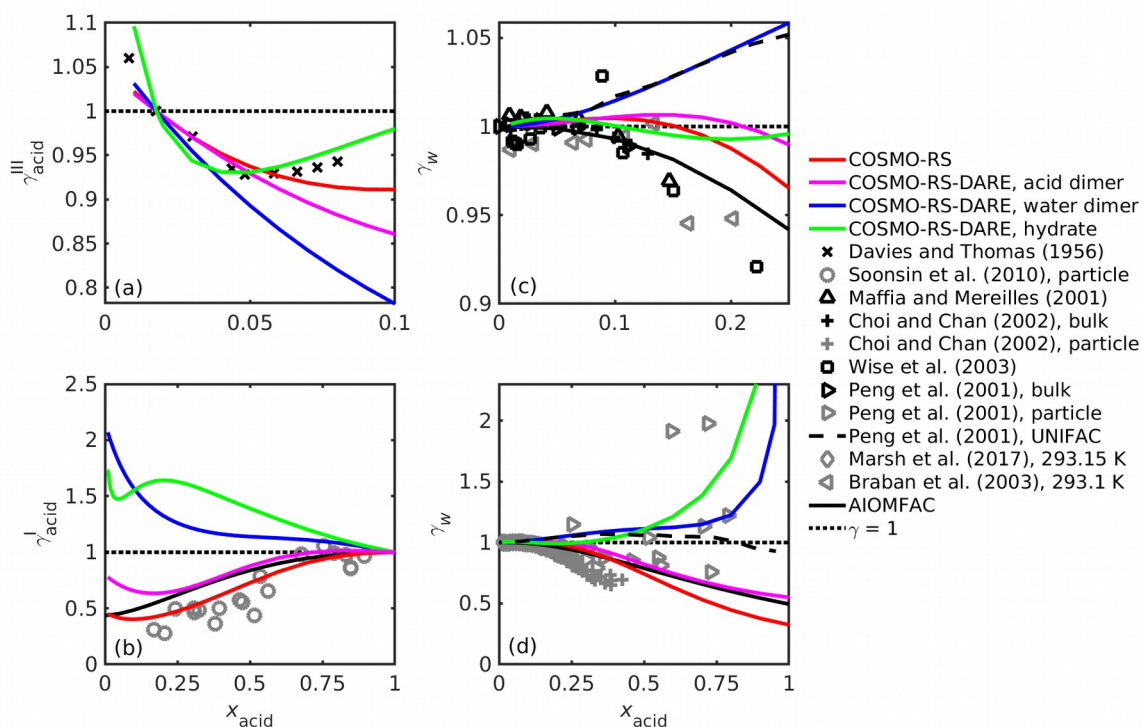
Author's response: We agree with the review's comment that some dicarboxylic acids have high solubilities. This section of the manuscript was reformulated to include comparison with additional experimental activity coefficients and UNIFAC predictions (see response to a comment from referee #1 on pages 3-6).

Line 247: Figs S2 and S3 should be moved to the main manuscript.

Line 255: Fig. S4 should be moved to the main manuscript.

Author's response: Figure S2 was moved to the main manuscript as an example of a comparison with experiments and UNIFAC (AIOMFAC). The comparisons with glutaric, oxalic, succinic, adipic and pimelic acid show similar agreement, and we therefore left Figs S3-S5 in the Supplement.

### Changes in manuscript:



**Figure 5. Activity coefficients of (a-b) malonic acid and (c-d) water in the binary mixtures at 298.15 K calculated using different clustering reactions in the COSMO-RS-DARE calculation. As a comparison are activity coefficients of malonic acid by Davies and Thomas (1956) (at 298.15 K given in convention III) and Soonsin et al. (2010) (particle measurements at various temperatures given in convention I) and of water by Maffia and Mereilles (2001), Choi et al. (2002), Wise et al. (2003), Peng et al. (2001), Marsh et al. (2017), Braban et al. (2003) and AIOMFAC-web (2020).**

Line 264: The logarithmic plot is not very informative. Rather show the figures from the SI here.

*Author's response: The x-scale of this figure was changed to linear and the whole mixing range was plotted in the figure. See response to the last comment of referee #1 (page 14-15).*

**New references:**

**AIOMFAC-web: version 2.32, <http://www.aiomfac.caltech.edu>, 2020.**

**Ben-Naim, A.: Solvation Thermodynamics, Plenum Press, New York and London, 1987.**

**Braban, C. F., Carroll, M. F., Styler, S. A., and Abbatt, J. P. D.: Phase transitions of malonic and oxalic acid aerosols, *J. Phys. Chem. A*, 107,6594-6602, <https://doi.org/10.1021/jp034483f>, 2003.**

**Eckert, F. and Klamt, A.: COSMOthermReference Manual, version C30, Release 19, COSMOlogic GmbH & Co, KG.: Leverkusen, Germany, 2019.**

**Fredenslund, A., Jones, R. L., and Prausnitz, J. M.: Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.*, 21, 1086-1099, <https://doi.org/10.1002/aic.690210607>, 1975.**

**Klamt, A. and Schüürmann, G.: COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin Trans. 2*, pp. 799-805, <https://doi.org/10.1039/P29930000799>, 1993.**

**Marsh, A., Miles, R. E. H., Rovelli, G., Cowling, A. G., Nandy, L., Dutcher, C. S., and Reid, J. P.: Influence of organic compound functionality on aerosol hygroscopicity: dicarboxylic acids, alkyl-substituents, sugars and amino acids, *Atmos. Chem. Phys.*, 17, 5583, <https://doi.org/10.5194/acp-17-5583-2017>, 2017.**

**Sordo, J. Á.: Solvation thermodynamics: two formulations and some misunderstandings, *RSC Adv.*, 5, 96 105-96 116, <https://doi.org/10.1039/c5ra17305a>, 2015.**

**Zuend, A., Marcolli, C., Luo, B. P., and Peter, T.: A thermodynamic model of mixed organic-inorganic aerosols to predict activity coefficients, *Atmos. Chem. Phys.*, 8, 4559-4593, <https://doi.org/10.5194/acp-8-4559-2008>, 2008.**

**Zuend, A., Marcolli, C., Booth, A. M., Lienhard, D. M., Soonsin, V., Krieger, U. K., Topping, D. O., McFiggans, G., Peter, T., and Seinfeld, J. H.: New and extended parameterization of the thermodynamic model AIOMFAC: calculation of activity coefficients for organic-inorganic mixtures containing carboxyl, hydroxyl, carbonyl, ether, ester, alkenyl, alkyl, and aromatic functional groups, *Atmos. Chem. Phys.*, **11**, 9155-9206, <https://doi.org/10.5194/acp-11-9155-2011>, 2011.**

# Technical note: Estimating aqueous solubilities and activity coefficients of mono- and $\alpha,\omega$ -dicarboxylic acids using **COSMO-RS-DARE** *COSMOtherm*

Noora Hyttinen<sup>1,2</sup>, Reyhaneh Heshmatnezhad<sup>1</sup>, Jonas Elm<sup>3</sup>, Theo Kurtén<sup>2</sup>, and Nønne L. Prisle<sup>1</sup>

<sup>1</sup>Nano and Molecular Systems Research Unit, University of Oulu, P.O. Box 3000, FI-90014 Oulu, Finland

<sup>2</sup>Department of Chemistry and Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, P.O. Box 55, FI-00014 Helsinki, Finland

<sup>3</sup>Department of Chemistry and iClimate, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark

**Correspondence:** Noora Hyttinen (noora.hyttinen@oulu.fi), Nønne L. Prisle (nonne.prisle@oulu.fi)

**Abstract.** We have used the *COSMOtherm* program to estimate activity coefficients and solubilities of mono- and  $\alpha,\omega$ -dicarboxylic acids, and water in binary acid–water systems. The deviation from ideality was found to be larger in the systems containing larger acids than in the systems containing smaller acids. [Conductor-like Screening Model for Real Solvents \(COSMO-RS\) underestimates experimental monocarboxylic acid activity coefficients by less than a factor of 2 but experimental water activity coefficients are underestimated more especially at high acid mole fractions.](#) We found a better agreement between [estimated \*COSMOtherm\*-estimated](#) and experimental activity coefficients of monocarboxylic acids when the water clustering with a carboxylic acid and itself was taken into account using the dimerization, aggregation and reaction extension (COSMO-RS-DARE) of *COSMOtherm*. [COSMO-RS-DARE is not fully predictive, but fitting parameters found here can be used to estimate thermodynamic properties of monocarboxylic acids in other aqueous solvents, such as salt solutions. For the dicarboxylic acids, COSMO-RS is sufficient for predicting aqueous solubility and activity coefficients and no fitting to experimental values is needed. This is highly beneficial for applications to atmospheric systems, as this data is typically not available for a wide range of mixing states realized in the atmosphere, either due to feasibility of the experiments or to sample availability.](#) Based on effective equilibrium constants of different clustering reactions in the binary solutions, acid dimer formation is more dominant in systems containing larger dicarboxylic acids (C5-C8), while for monocarboxylic acids (C1-C6) and smaller dicarboxylic acids (C2-C4), hydrate formation is more favorable, especially in dilute solutions.

## 1 Introduction

Mono- and dicarboxylic acids ( $\text{CH}_3(\text{CH}_2)_{n-2}\text{COOH}$  and  $\text{COOH}(\text{CH}_2)_{m-2}\text{COOH}$ , respectively) are common atmospheric compounds that have been detected in both the gas (Kawamura et al., 2000; Fisseha et al., 2006) and aerosol phase (Fisseha et al., 2006; Verma et al., 2017; Guo et al., 2014; Hyder et al., 2012). Carboxylic acids have been detected in high abundance in various environments, such as urban (Zhao et al., 2018; Kawamura et al., 2000; Fisseha et al., 2006; Jung et al., 2010; Guo et al., 2014), semi-urban (Verma et al., 2017), marine (Kawamura and Sakaguchi, 1999; Mochida et al., 2003) and Antarctic (Kawamura et al., 1996) measurement sites. In general, small carboxylic acids ( $n \leq 3$  and  $m \leq 4$ ) are more abundant than

large acids ( $n > 3$  and  $m > 4$ ) (Jung et al., 2010; Fisseha et al., 2006; Tsai and Kuo, 2013; Zhao et al., 2018; Guo et al., 2014; Kawamura et al., 2000). For example, Tsai and Kuo (2013) found that 77.2% of all carboxylic acids in fine particulate matter (PM<sub>2.5</sub>) were small carboxylic acids (formic, acetic and oxalic acid) in a broad-leaved forest in central Taiwan.

Accurate description of the different aerosol phases is important for determining parameters used in aerosol modeling, such as gas-to-particle partitioning, in particular water uptake and chemical reactivity. A large number of reactions in the aqueous aerosol phase are strongly pH dependent (Pye et al., 2020; Weber et al., 2016), but accurate predictions of aerosol acidity are highly challenging. One element to resolve is the nature and amount of acidic material dissolved in the aqueous aerosol phase. The aqueous bulk solubility of mono- and dicarboxylic acids have been measured in multiple studies (Saxena and Hildemann, 1996; Apelblat and Manzurola, 1987, 1989, 1990; Cornils and Lappe, 2000; Song et al., 2012; Romero and Suárez, 2009; Omar and Ulrich, 2006; Brooks et al., 2002). However, ~~the acid~~ activity data of carboxylic acid–water systems is much ~~more scarce~~scarcer. Jones and Bury (1927) derived the activity coefficients of formic ( $n = 1$ ), acetic ( $n = 2$ ), propanoic ( $n = 3$ ) and butanoic ( $n = 4$ ) acids in aqueous solutions at the freezing points of the binary solutions using freezing point depression measurements. Using freezing point depression measurements, activity coefficients are calculated using Lewis and Randall’s equation for non-electrolytes. Hansen et al. (1955) derived activity coefficients of acetic, propanoic and butanoic acid in water and the activity coefficients of water in acetic, propanoic, butanoic, pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acids, at 298.15 K, using partial pressure measurements. In addition, ~~the Gibbs–Duhem equation was fitted to the experimental activity coefficients of the acetic, propanoic and butanoic acid–water systems. Most of the activity data of aqueous dicarboxylic acid systems in the literature (Davies and Thomas, 1956; Maffia and Meirelles, 2001; Peng et al., 2001; Choi and Chan, 2002; Wise et al., 2003) is only obtained from measurements on dilute solutions due to the low solubility of the acids in water.~~ Hansen et al. (1955) represented the experimental points using self-consistent activity coefficient functions. Activity coefficients of malonic, succinic and glutaric acid ( $m = 3, 4$  and  $5$ ) have been measured by Davies and Thomas (1956) and Soonsin et al. (2010) in bulk and particle experiments, respectively.

~~The solubilities~~ Group contribution methods, such as UNIFAC (Fredenslund et al., 1975) and AIOMFAC (Zuend et al., 2008), are often used to estimate activity coefficients of atmospherically relevant compounds. More recently, a quantum chemistry based Conductor-like Screening Model for Real Solvents (COSMO-RS; Klamt, 1995; Klamt et al., 1998; Eckert and Klamt, 2002) has been used to predict thermodynamic properties of multifunctional compounds. Solubilities and activity coefficients of carboxylic acids have previously been estimated using the COSMO-RS theory (~~Klamt, 1995; Klamt et al., 1998; Eckert and Klamt, 2002~~) implemented in the COSMO~~therm~~ program (COSMO~~therm~~, 2019). Schröder et al. (2010) estimated the aqueous solubilities of various polycarboxylic acids using the TZVP parametrization of COSMO~~therm~~ and found that COSMO~~therm~~ was able to predict the temperature dependence of the solubilities of dicarboxylic acids ( $m = 2–8$ ) well, while the absolute solubility estimates were not in a good agreement with experiments. Michailoudi et al. (2020) estimated the activity coefficients of monocarboxylic acids with even number of carbon atoms ( $n = 2, 4, 6, 8, 10, 12$ ) in the infinite dilution state. In addition, they estimated the solubility of the same acids in pure water and different aqueous electrolyte solutions. They found a good agreement between experimental and estimated aqueous solubilities of the acids with the exception of butanoic acid, which in experiments has been seen to be fully soluble (Saxena and Hildemann, 1996), while COSMO~~therm~~ predicted a finite solubility.



Recent work has shown that the absolute COSMO $therm$  solubility and activity coefficient estimates can be improved by excluding conformers containing intramolecular hydrogen bonds from the COSMO $therm$  calculation (Hytinen and Prisle, 2020). However, based on the hydrogen bonding definition of COSMO $therm$ , monocarboxylic acids are not able to form intramolecular hydrogen bonds. Therefore, other methods are needed to improve COSMO $therm$  estimates of monocarboxylic acids. On the other hand, carboxylic acids are able to form hydrogen bonded dimers where two molecules are bound by two simultaneous intermolecular hydrogen bonds. These concerted multiple contacts, such as is seen in carboxylic acid dimer formation, are not captured by COSMO-RS. A dimerization, aggregation and reaction extension to the COSMO-RS theory (COSMO-RS-DARE) was developed to account for these interactions (Sachsenhauser et al., 2014). For example, Cysewski (2019) was able to improve the agreement between experimental and estimated solubilities of ethenzamide in various organic solvents using COSMO-RS-DARE.

~~Here we~~ Most atmospherically relevant multifunctional compounds are not readily available for experimental determination of thermodynamic properties. Accurate theoretical estimates are therefore essential for advancing current aerosol process modeling to include more complex compounds and mixtures. Here, we demonstrate the applicability of COSMO-RS theory in calculating condensed-phase properties of atmospherically relevant organic compounds. Carboxylic acids are among the most abundant and well characterized organic compounds in the troposphere and are therefore a good compound class to use to validate the use of COSMO-RS in atmospheric research. We use the newly developed COSMO-RS-DARE~~method~~, as well as COSMO-RS, to estimate activity coefficients of monocarboxylic acids ( $n = 1-6$ ) and  $\alpha,\omega$ -dicarboxylic acids ( $m = 2-8$ ), and water, in binary acid-water mixtures. In addition, we estimate aqueous solubilities and effective equilibrium constants of cluster formation of the acids using both the COSMO-RS-DARE and COSMO-RS methods.

## 2 COSMO $therm$ calculations

We use the COSMO $therm$  software (release 19 and parametrization BP\_TZVPD\_FINE\_19) (COSMO $therm$ , 2019) to estimate the solubilities and activity coefficient of linear mono- and dicarboxylic acids in binary aqueous solutions. In addition, we compute the effective equilibrium constants of water and acid dimerization (formation of a hydrogen bonded cluster containing two water molecules or two acid molecules, respectively), and acid hydration (formation of a hydrogen bonded cluster containing one acid and one water molecule).

### 2.1 Activity coefficients

COSMO $therm$  calculates the activity coefficient ( $\gamma$ ) of compound  $i$  with mole fraction  $x_i$  using the pseudo-chemical potentials at composition  $\{x_i\}$  ( $\mu_i^*(x_i)$ ) and at the reference state ( $\mu_i^{*\circ}(x^\circ, T, P)$ ). By default, the reference state used in COSMO $therm$  is the pure compound (labeled as convention I (Levine, 2009); Levine, 2009):

$$\ln \gamma_i^I(x_i) = \frac{\mu_i^*(x_i) - \mu_i^{*\circ, I}(x^\circ, T, P)}{RT} \quad (1)$$

at  $P = 10^5$  Pa reference pressure.  $T$  is the temperature (K) and  $R$  the gas constant ( $\text{kJ K}^{-1} \text{mol}^{-1}$ , when  $\mu^*$  is given in  $\text{kJ mol}^{-1}$ ).

- 90 ~~The pseudo-chemical potential is~~ Pseudo-chemical potential (Ben-Naim, 1987) is an auxiliary quantity defined using the standard chemical potential at the reference state  $\mu^\circ$ :

$$\mu_i^*(x_i) = \mu_i^\circ(x^\circ, T, P) + RT \ln \gamma_i(x_i) \quad (2)$$

at  $P = 10^5$  Pa reference pressure.  $T$  is the temperature. Pseudo-chemical potential has recently been used in molecular level solvation thermodynamics as a replacement to chemical potential (Sordo, 2015). By definition, the activity coefficient of a compound at the reference state is unity ( $\gamma_i^I(x_i = 1) = 1$ ), which leads to equal chemical and pseudo-chemical potential at the reference state. At other states ( $x_i < 1$ ), the relation between chemical and pseudo-chemical potentials ( $\mu$  and  $R$  the gas constant ( $\text{kJ K}^{-1} \text{mol}^{-1}$ ))  $\mu^*$ , respectively) can be expressed as

- 95

$$\mu_i^*(x_i) = \mu_i(x_i) - RT \ln x_i \quad (3)$$

100 Unless otherwise mentioned, the mole fractions  $x_i$  correspond to mole fractions of undissociated acid or neutral non-protonated water.

## 2.2 Solubility

Solubilities are calculated by finding the liquid-liquid (LLE) or the solid-liquid equilibrium (SLE) of the binary liquid–water and/or solid–water systems, respectively. ~~The LLE is found by solving the~~ In LLE, the standard chemical potential ( $\mu$ ) of a compound is equal in both of the liquid phases ( $\alpha$  and  $\beta$ ):

- 105 
$$\mu_i(x_i^\alpha) = \mu_i(x_i^\beta) \quad (4)$$

The standard chemical potential of compound  $i$  in a solution is defined using the standard chemical potential at the reference state:

$$\mu_i(x_i) = \mu_i^\circ(x^\circ, T, P) + RT \ln a_i(x_i), \quad (5)$$

- 110 where  $a_i(x_i)$  is the activity of compound  $i$  with mole fraction  $x_i$ . The liquid-liquid equilibrium condition between the solvent rich solvent-rich phase ( $\alpha$ ) and the solute rich solute-rich phase ( $\beta$ ) becomes:

$$a_i(x_i^\alpha) = a_i(x_i^\beta) \quad (6)$$

At the LLE, the activity ( $a_i^I(x_i) = \gamma_i^I(x_i)x_i$ ) of both the solvent and the solute is equal in the two phases. For example, for the acid solute:-

$$a_i^{I,\alpha}(x_{\text{SOL,acid}}) = a_i^{I,\beta}(1 - x_{\text{SOL,w}})$$

115 The SLE is solved from the solid-liquid equilibrium condition ~~÷(Eckert and Klamt, 2019):~~

$$\log_{10}(x_{\text{SOL},i}) = \frac{\mu_i^{*\circ,I}(x^\circ, T, P) - \mu_i^*(x_{\text{SOL},i}) - \Delta G_{\text{fus}}(T)}{RT \ln(10)}, \quad (7)$$

~~where  $x_{\text{SOL},i}$  is the mole fraction solubility (SOL) of compound  $i$  in the solvent.~~ The free energy of fusion of the solute ( $\Delta G_{\text{fus}}(T)$ ) is calculated from the experimentally determined heat of fusion ( $\Delta H_{\text{fus}}$ ) and melting ~~temperature point~~ ( $T_{\text{melt}}$ ) using the Schröder-van Laar equation (Prigogine and Defay, 1954):

$$120 \quad \Delta G_{\text{fus}}(T) = \Delta H_{\text{fus}} \left(1 - \frac{T}{T_{\text{melt}}}\right) - \Delta C_{p,\text{fus}}(T_{\text{melt}} - T) + \Delta C_{p,\text{fus}} T \ln \frac{T_{\text{melt}}}{T} \quad (8)$$

Here the heat capacity of fusion ( $\Delta C_{p,\text{fus}}$ ) is estimated from the melting point and the heat of fusion:

$$\Delta C_{p,\text{fus}} = \frac{\Delta H_{\text{fus}}}{T_{\text{melt}}} \quad (9)$$

Table 1 shows ~~the literature values of experimental~~ melting points and heats of fusion of the dicarboxylic acids of this study. Melting points and heats of fusion of the monocarboxylic acids are not used, since all of the monocarboxylic acids studied here  
125 are in liquid phase at 298.15 K.

### 2.3 ~~Equilibrium-Effective equilibrium~~ constants

COSMOtherm estimates ~~effective~~ equilibrium constants of condensed-phase reactions from the free energy of the reaction ( ~~$\Delta G_r$~~ :  ~~$\Delta G_r^{\text{Io}}$~~ ):

$$K_{\text{r,eff}} = e^{-\frac{\Delta G_r}{RT} - \frac{\Delta G_r^{\text{Io}}}{RT}} \quad (10)$$

130 The reaction free energy is calculated from the free energies of the ~~reactants~~ ( ~~$G_{\text{react}}$~~  ~~pure reactants~~ ( ~~$G_{\text{react}}^{\text{Io}}$~~ ) and products ( ~~$G_{\text{prod}}$~~ ) ~~in the solution~~:  ~~$G_{\text{prod}}^{\text{Io}}$~~ ):

$$\Delta G_r \Delta G_r^{\text{Io}} = \Sigma G_{\text{prod}}^{\text{Io}} - \Sigma G_{\text{react}}^{\text{Io}} \quad (11)$$

The free energy of compound  $i$  (~~reactant or product~~) is the sum of the energy of the solvated compound ( $E_{\text{COSMO}}$ ), the averaged correction for the dielectric energy ( $dE$ ; Klamt et al., 1998) and the ~~chemical-pseudo-chemical~~ potential of the ~~compound in~~  
135 ~~the solution~~:

$$G_i = E_{\text{COSMO},i} + dE_i + \mu_i^*(x_i)$$

~~The solvent-independent effective equilibrium constant can be calculated by multiplying the reaction equilibrium constant with the ratio of the activity coefficients of the reactants ( $\gamma_{\text{react}}$ ) and products ( $\gamma_{\text{prod}}$ ):~~

$$K_{\text{eff}} = \frac{\Pi \gamma_{\text{prod}}}{\Pi \gamma_{\text{react}}} K_r$$

140 Combining Eqs ?? and ?? gives:-

$$\ln K_{\text{eff}} = \Sigma(\ln \gamma_{\text{prod}}) - \Sigma(\ln \gamma_{\text{react}}) - \frac{\Sigma G_{\text{prod}} - \Sigma G_{\text{react}}}{RT}$$

where the activity coefficients can be replaced by Eq. 1 and the free energies by Eq. ??:-

$$RT \ln K_{\text{eff}} = \Sigma(\mu_{\text{prod}}^*(x_{\text{prod}}) - \mu_{\text{prod}}^{*\circ, \text{I}}(x^\circ, T, P)) - \Sigma(\mu_{\text{react}}^*(x_{\text{react}}) - \mu_{\text{react}}^{*\circ, \text{I}}(x^\circ, T, P))$$

145

$$\begin{aligned} & -\Sigma(E_{\text{COSMO, prod}} + dE_{\text{prod}} + \mu_{\text{prod}}^*(x_{\text{prod}})) + \Sigma(E_{\text{COSMO, react}} + dE_{\text{react}} + \mu_{\text{react}}^*(x_{\text{react}})) \\ & = -\Sigma(E_{\text{COSMO, prod}} + dE_{\text{prod}} + \mu_{\text{prod}}^{*\circ, \text{I}}(x^\circ, T, P)) + \Sigma(E_{\text{COSMO, react}} + dE_{\text{react}} + \mu_{\text{react}}^{*\circ, \text{I}}(x^\circ, T, P)) \\ & = -\Sigma G_{\text{prod}}^{\circ, \text{I}} + \Sigma G_{\text{react}}^{\circ, \text{I}} \end{aligned}$$

150 This gives us the final equation for calculating the effective equilibrium constant of a reaction: pure compound:

$$K_{\text{eff}} G_i^{\text{Io}} = e^{-\frac{\Delta G_r^{\circ, \text{I}}}{RT}} E_{\text{COSMO, } i} + dE_i + \mu_i^{*\circ, \text{Io}}(x^\circ, T, P) \quad (12)$$

where the equilibrium constant depends on only the free energies of the pure compounds ( $G_i^{\circ, \text{I}}$ ). COSMOtherm calculates the equilibrium constants on a mole fraction basis, which leads to dimensionless equilibrium constants:-

## 2.4 Concentration dependent reactions (COSMO-RS-DARE)

155 In COSMO-RS, the surface of a molecule is divided into surface segments that represent the surface charges of the molecule. The surface is considered as an interface between a virtual conductor around the molecule and the cavity formed by the molecule (Klamt and Schüürmann, 1993). Each surface segment has an area ( $[\text{\AA}^{-2}]$ ) and a screening charge density ( $\sigma$  [ $e \text{\AA}^{-2}$ ]). Interactions between molecules are described through the interaction between surface segments of the molecules. different molecules. Examples of  $\sigma$ -surfaces used in COSMOtherm calculations are shown in Fig. 1. The red color of a  
160  $\sigma$ -surface signifies a positive screening charge density (negative partial charge) and the blue color a negative screening charge density (positive partial charge).

Concerted multiple contacts, such as carboxylic acid dimer formation, are not captured by COSMO-RS. COSMOtherm is able to consider these hydrogen bonded clusters using the dimerization, aggregation, and reaction extension (COSMO-RS-DARE, Sachsenhauser et al. (2014)). We use the COSMO-RS-DARE method in our activity coefficient and solubility calculations. In our equilibrium constant calculations, the clusters in the system are included as the product of the clustering reactions. The method is described below and the full COSMO-RS-DARE derivation can be found from Sachsenhauser et al. (2014).

A clustering reaction between molecules A and B can be described by the equilibrium:



In acid-water systems, A and B can be either a carboxylic acid or a water molecule. In COSMO-RS-DARE, the product clusters are included in property-COSMOtherm calculations by using the screening charge density surface ( $\sigma$ -surface)-surfaces of

molecule A in the cluster, omitting the part of the  $\sigma$ -surface that is assigned to the molecule clustered with A (i.e., molecule B). Similarly, the clustering product of molecule B is included in the calculation by omitting the  $\sigma$ -surface assigned to molecule A from the  $\sigma$ -surface of A·B. Examples of these partial  $\sigma$ -surfaces are shown on the right hand side of Fig. 1.

The formation of hydrogen bonds (in hydrates or dimers) is taken into account using the interaction energy of the two reacting compounds. The free energy between the interaction sites formation free energy of the cluster ( $G(A, A \cdot B)$ ) is calculated using the energy difference between the monomer compound (A) and the same compound fitting parameters  $c_H$  and  $c_S$  (enthalpic and entropic contributions, respectively) to describe the interaction between the monomers A and B in the cluster (A·B),  $\Delta^0$ :

$$G(A, A \cdot B) = -2\Delta^0 + c_H - c_S T, \quad (13)$$

where fitting parameters  $c_H$  and  $c_S$  are the enthalpic and entropic contributions to the interaction energy, respectively. Because we are only calculating condensed-phase properties at one temperature ( $T = 298.15$  K), we are only fitting. The fitting parameters are used because COSMOtherm is unable to calculate the energy of a monomer in a cluster. Instead, the energy of a monomer in a cluster is assumed to be equal to the energy of the lowest-energy conformer of the same compound and the favorability of the cluster formation is estimated using the fitting parameters. Without temperature dependent experimental data, it is not possible to fit both fitting parameters. We therefore consider the enthalpic parameter  $c_H$ , keeping as the total formation free energy parameter at 298.15 K, setting the entropic parameter  $c_S$  at to zero.

COSMO-RS-DARE was originally developed for systems containing carboxylic acids in non-polar solvents (Sachsenhauser et al., 2014). In a carboxylic acid–water system, both the carboxylic acid and water are able to form strongly bound clusters. In addition, hydrated acids can be formed. We are thus including the interactions of the clustering reactions for both A and B, even when A = B.

## 190 2.5 Input file generation

The .cosmo files of water and the monocarboxylic acids with a low number of conformers ( $<10$ ) are taken from the COSMObase (COSMObase, 2011) database. For the dicarboxylic acids, acid and water dimers, and the hydrates of pimelic ( $m = 7$ ) and suberic ( $m = 8$ ) acids, we use the following systematic conformer search approach detailed by Kurtén et al. (2018) as it has been shown to give more consistent results than other conformer sampling approaches. The conformers are found using the systematic conformer search in the Spartan program (Wavefunction Inc., 2014, 2016). The conformer set is then used as input to the COSMOconf program (COSMOconf, 2013) (using the TURBOMOLE program (TURBOMOLE, 2010)), which runs initial single-point COSMO calculations at the BP/def-SV(P) level of theory to compare the chemical-pseudo-chemical potentials of the conformers and remove similar structures. Initial geometry optimizations are calculated at the BP/def-SV(P) level of theory, duplicate structures are removed by comparing the new geometries and chemical-pseudo-chemical potentials. Final geometries are optimized at the BP/def-TZVP level of theory and after a second duplicate removal step, final single-point energies are calculated at the BP/def2-TZVPD-FINE level of theory.

For acid dimers, we use the lowest gas-phase energy structures found by Elm et al. (2019) as a starting structure for systematic conformer search. For hydrated monocarboxylic acids and smaller dicarboxylic acids ( $m \leq 6$ ), the clusters are built

by adding a water molecule to each conformer of the free acids. For monocarboxylic acids, the water molecule is placed on the carboxylic acid group forming two intermolecular hydrogen bonds between the molecules. For dicarboxylic acids, a water molecule is added to either end of the acid, forming two hydrate conformers from a single acid conformer. For the dicarboxylic acid conformers with the two acid groups close to each other, additional conformers are created for cases where the water molecule is interacting with both acid groups. [Fig. 2 illustrates the formation of two different adipic acid hydrate conformers from a single monomer conformer. A cluster conformer where the water molecule is attached to one carboxylic acid group is shown on the top right corner and in the bottom right corner conformer, the water molecule is bound to both acid groups.](#) Due to the large number of conformers of non-hydrated pimelic ( $m = 7$ ) and suberic ( $m = 8$ ) acid (75 and 132, respectively), the monohydrate conformers of those two acids are sampled separately using Spartan.

We use only clusters of two molecules in our calculations. In carboxylic acid dimers, the hydrogen bond donors and acceptors are saturated, which means that carboxylic acids are unlikely to form larger clusters than dimers (Vawdrey et al., 2004; Elm et al., 2014, 2019). Computational studies (Aloisio et al., 2002; Weber et al., 2012; Kildgaard et al., 2018) have shown that, in the gas phase, the energetically most favorable dihydrate is formed by two water molecules attaching to the same carboxylic acid group. Therefore, adding a second water molecule to the cluster does not significantly change the probability distribution of the screening charge density ( $\sigma$ -profile) of the acid in the cluster compared to the acid in a monohydrate or dimer.

Conformers containing no intramolecular hydrogen bonds (Kurtén et al., 2018; Hyttinen and Prisle, 2020) are used in the COSMO-RS solubility and activity coefficient calculations. Due to the intermolecular hydrogen bonding in the hydrate and dimer clusters, all conformers (up to 40 conformers) of monomers and clusters are used in the effective equilibrium constant calculations. In COSMO-RS-DARE calculations, we use all conformers of the monomers and only the lowest solvated energy conformers of the clusters.

### 3 Results and discussion

#### 3.1 Effective equilibrium constants of clustering reactions

We estimated the effective equilibrium constants of the different clustering reactions (i.e., hydration and dimerization) of the binary acid–water systems. A comparison between the hydration and acid dimerization equilibrium constants in the aqueous phase is given in Fig. 3 and Table 2. The equilibrium constants for both the dimerization and hydration reactions are similar between all of the monocarboxylic acids, and are not labeled in Fig. 3. For the dicarboxylic acid, we can see larger variation in both the hydration and dimerization reactions. Note that the COSMO-RS-DARE method is not used in the effective equilibrium constant calculations because the clusters are already included in the calculation as products.

For all of the acids, the effective equilibrium constant of dimerization is higher than that of the hydrate formation of the corresponding acid, meaning that acid dimer formation is energetically more favorable than hydrate formation. However, in dilute conditions, water is more abundant, shifting the equilibrium from acid dimerization to hydration. The dimerization:hydration ratio is the lowest for oxalic ( $m = 2$ ) and malonic ( $m = 3$ ) acids, while monocarboxylic acids and succinic acid ( $m = 4$ )

have similar (intermediate) ratios, and the larger dicarboxylic acids ( $m = 5-8$ ) have higher ratios. This means that, in dilute solutions, oxalic, malonic and succinic acid will most likely interact with water instead of other acid molecules.

Vawdrey et al. (2004) calculated the dimerization enthalpies (at the B3LYP/6-31++G(2d,p) level of theory) of monocarboxylic acids ( $n = 2-6$ ) and found a notable even-odd variation (dimerization of the acids with odd number of carbon atoms is more favorable than of the even carbon number acids). The same is seen here in the condensed phase, where the effective equilibrium constant of butanoic and hexanoic acids are lower than of propanoic and pentanoic acids, respectively. Otherwise there is a slightly increasing trend in the effective equilibrium constants with the increasing number of carbon atoms in the monocarboxylic acids. For larger dicarboxylic acids ( $m \geq 4$ ), Elm et al. (2019) found an even-odd alternation in the dimer:monomer ratio in the gas phase, calculated at the DLPNO-CCSD(T)/aug-cc-pVTZ// $\omega$ B97X-D/6-31++G(d,p) level of theory. We observe a similar increasing effective equilibrium constant with the increasing carbon chain length in the smaller dicarboxylic acids ( $m = 2-5$ ) and an even-odd alternation in the larger dicarboxylic acids ( $m = 4-8$ ) in any condensed phase.

## 3.2 Activity coefficients

### 3.2.1 Monocarboxylic acids

We calculated the activity coefficient of the carboxylic acids and water in the binary acid-water mixtures using the COSMO-RS-DARE method. Hansen et al. (1955) derived the activity coefficients of acetic ( $n = 2$ ), propanoic ( $n = 3$ ) and butanoic ( $n = 4$ ) acids in mixtures with water from partial pressure measurements. In addition, they determined activity coefficients of water in aqueous acetic, propanoic, butanoic, pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acid mixtures. We used these experimental activity coefficients to fit the enthalpic parameters ( $c_H$ ) for each of the acids in the COSMO-RS-DARE calculations. Figure 4 shows a comparison between the estimated and experimentally determined activity coefficients of these monocarboxylic acids, and formic acid ( $n = 1$ ), for which no experimental activity coefficient data is available.

The reactions included in the calculations are water dimer ( $\text{H}_2\text{O} \cdot \text{H}_2\text{O}$ ) and acid hydrate ( $\text{RCOOH} \cdot \text{H}_2\text{O}$ ) formation. A comparison between COSMO-RS activity coefficients and the COSMO-RS-DARE method with different clusters included in the calculation, are shown for acetic acid in Fig. S1 of the Supplement. For acetic acid, we found the best fit between experimental and estimated activity coefficients using  $c_H = 0 \text{ kcal mol}^{-1} \text{ kJ mol}^{-1}$  for both the water dimerization and acid hydration reactions. Decreasing the  $c_H$  of either clustering reaction leads to stronger deviation from ideality, which in our case leads to a worse fit for water activity coefficient, and positive parameter values cannot be used to lower the interaction enthalpy. The effective equilibrium constant for water dimer formation ( $5.71 \times 10^5$ ) is below that of acetic acid hydration ( $4.36 \times 10^6$ ), which explains why the fitting parameter of the water dimer hydration should be higher (or equal, since positive values are not possible) than the parameter for acid hydrate formation. We additionally calculated UNIFAC predictions of acid and water activity coefficients using AIOMFAC-web (AIOMFAC-web, 2020; Zuend et al., 2008, 2011). These calculations (without inorganic ions) correspond to modified UNIFAC calculations by Peng et al. (2001). From Fig. S1 we see that, for acetic acid, the UNIFAC model underestimates the experimental activity coefficients more than even the COSMO-RS estimate.

Similar to COSMO-RS, UNIFAC is not able to predict the increasing trend of water activity coefficients with the increasing acid mole fraction.

270 For the other monocarboxylic acids studied here, we used the same  $c_H$  value for water dimerization that was found for the acetic acid–water system, and fitted the  $c_H$  of acid hydrate formation to the experimental activity coefficients of water and the acids in the corresponding acid–water systems (Hansen et al., 1955). The enthalpic parameter values of acid hydration used to estimate the activity coefficients shown in Fig. 4 are 0.0, 0.0, ~~-2.5, -3.5, -2.2 and -2.0 kcal mol<sup>-1</sup>~~ -10.5, -14.6, -9.2 and -8.4 kJ mol<sup>-1</sup>, for formic ( $n = 1$ ), acetic ( $n = 2$ ), propanoic ( $n = 3$ ), butanoic ( $n = 4$ ), pentanoic ( $n = 5$ ) and hexanoic ( $n =$   
275 6) acid, respectively. For formic acid, we used the same  $c_H$  parameter as for acetic acid due to lack of experimental activity coefficients. If the enthalpic parameters in COSMO-RS-DARE calculations are not fitted to experimental activity coefficients and instead are set to zero, the activity coefficients of both acid and water underestimate the experimental activity coefficients of Hansen et al. (1955) (see Fig. S2 of the Supplement). If no experimental activity coefficients are available for fitting the COSMO-RS-DARE parameters, COSMO-RS estimates agree with experiments overall better than COSMO-RS-DARE or  
280 UNIFAC. COSMO-RS-estimated acid activity coefficients are close to the measured values in all mixing states, and for water activity coefficients the agreement between COSMO-RS and experiments is good in mixing states with  $x_{\text{acid}} < 0.75$ .

Sachsenhauser et al. (2014) used the COSMO-RS-DARE method for binary systems containing either acetic ( $n = 2$ ) or propanoic ( $n = 3$ ) acid and a non-polar organic solvent. Their calculations show that the dimerization parameter (equivalent to  $c_H$  in our calculations) is higher for propanoic acid than for acetic acid. This is opposite to what we observed for the hydration  
285 parameters, where  $c_H$  was found to be higher for acetic acid than for propanoic acid. This indicates that the fitting parameters of one clustering reaction cannot be used to estimate the corresponding fitting parameters of another clustering reactions of the same compound.

While COSMO-RS is fully predictive, COSMO-RS-DARE requires parameter fitting using experimental data. Fitted COSMO-RS-DARE parameters from one system can be used in other systems where the same clustering reactions are relevant. For instance,  
290 Sachsenhauser et al. (2014) found that the same interaction parameters of acid dimers can be used in systems containing other similar (non-polar) solvents. This indicates that our interaction enthalpies can be applied to other aqueous systems, for instance, ternary systems containing an inorganic salt, in addition to the carboxylic acid and water. This would allow for extending the findings of this study to atmospherically relevant aerosol solutions.

The increasing length of the acid carbon backbone leads to larger deviation from ideality ( $\gamma = 1$ ) for both the acid and water.  
295 In convention I, this means the acid and water activity coefficient values are higher in mixtures containing the longer acids **then** than the shorter acids. We observe that COSMO-RS-DARE estimated activity coefficients agree well with the experiments once the  $c_H$  parameter is fitted. However, when the hydrate and water dimer reactions are included, COSMO-RS-DARE is not able to predict realistic activity coefficients for water at high mole fractions ( $x_{\text{acid}} > 0.9$ ) of the acids. This is likely due to the low concentration of water in the binary solution, leading to errors in the description of the interactions between water molecules.  
300 **The Still**, COSMO-RS-DARE estimates agree well with the experiments at least up to 0.9 mole fraction of the monocarboxylic acids. **For higher** This is an improvement compared to the UNIFAC model, which fails to reproduce experimental water activity



coefficients already at acid mole fractions above 0.25. At very high acid mole fractions ( $x_{\text{acid}} > 0.95$ ), COSMO-RS-DARE predicts several orders of magnitude higher activity coefficients than what was seen in experiments.

### 3.2.2 Dicarboxylic acids

305 Multiple experimental studies have reported activity (coefficient) data for either dicarboxylic acids or water in binary acid–water systems (Davies and Thomas, 1956; Soonsin et al., 2010; Maffia and Meirelles, 2001; Peng et al., 2001; Choi and Chan, 2002; Wise et al., 2002). However, due to the low solubility of the dicarboxylic acids in water (see Fig. 8 for the experimental solubilities), fitting of the  $c_H$  parameter is not feasible for all of the acids. The concentration of the dicarboxylic acid at the reference state, with respect to which the experimental activity coefficients of dicarboxylic acids are determined, is either 0.5 or 1 mol kg<sup>-1</sup>, and the activity coefficients have only been measured in states close to the reference state (Davies and Thomas, 1956). In addition, Soonsin et al. (2010) reported activity coefficients for malonic, succinic and glutaric acid in convention I at various temperatures in the particle phase. However, Hyttinen and Prisle (2020) have previously found a good agreement between experimental particle phase activity coefficients and COSMO $_{therm}$  estimates of citric, tartaric, malic and malonic acid.

We tested the effect of including different clusters in the activity coefficient calculation of malonic acid ( $m = 3$ ) and glutaric acid ( $m = 5$ ). Both of these acids have relatively high aqueous solubilities ( $x_{\text{SOL,acid}} \approx 0.5$ ) and multiple experimental studies that have measured water activities in binary acid–water solutions. For the other dicarboxylic acids in this study, we were only able to find A comparison between the experimental, UNIFAC-modeled and COSMO $_{therm}$ -estimated activity coefficients are shown in Fig. 5. The malonic acid activity coefficients for succinic acid ( $m = 4$ ), at 298.15 K and the values were only measured in a small acid mole fraction range ( $0.007 < x_{\text{acid}} < 0.01$ ) with respect to  $x_{\text{acid}} = 0.009$ . A comparison between the experimentally determined activity coefficients and are compared in convention III (Fig. 5a) and in convention I (Fig. 5b). The COSMO $_{therm}$  estimates of malonic and glutaric acid are shown in Figs S2 and S3 of the Supplement, respectively. The agreement between experimental and estimated activity coefficients is not as good as for the monocarboxylic acid but the agreement is reasonable for both acids using the COSMO-RS method with conformers containing no intramolecular hydrogen bonds. For glutaric acid, using COSMO-RS-DARE with acid dimerization gives similar results as COSMO-RS (see Fig. S3) -estimated water activity coefficients are compared with experimental bulk (Fig. 5c) and particle (Fig. 5d) phase activity coefficients and UNIFAC-estimated activity coefficients.

For malonic acid (and other studied dicarboxylic acids, see Figs S3-S5 of the Supplement). For succinic acid, there is a large variation in the water activity coefficients between the different experiments and the COSMO-RS, COSMO-RS-DARE is not able to improve the agreement between experiments and COSMO-RS-DARE methods both agree reasonably well with the activity coefficients measured by Maffia and Meirelles (2001) and Peng et al. (2001). However, succinic acid activity coefficient estimates agree better with the particle phase activity coefficients of Soonsin et al. (2010) COSMO $_{therm}$  estimates, the best overall fit is found using COSMO-RS. The water activity coefficients estimated using COSMO-RS as opposed to COSMO-RS-DARE (Fig. S4 of the Supplement).

335 From Fig. 3, we can see that the larger dicarboxylic ( $m > 4$ ) acids have higher dimerization:hydration ratios than the monocarboxylic acids. Since glutaric acid ( $m = 5$ ) is better described by including the acid dimerization reaction to the system

than by including water dimerization and/or acid hydration, we used the acid dimerization ( $e_H = 0 \text{ kcal mol}^{-1}$ ) in the acid description of the other larger dicarboxylic acids ( $m = 6-8$ ). With relatively high effective hydration equilibrium constant, the hydrated oxalic acid is likely the most appropriate clustering reaction to be added in the COSMO-RS-DARE calculation of oxalic acid are close to ones estimated using the UNIFAC model (modified UNIFAC; Peng et al., 2001). However, without experimental activity coefficient data of oxalic acid-water systems, we are not able to fit the  $e_H$  parameter needed for reliable estimates. Similarly to what has been seen with the UNIFAC model, COSMOtherm is able to predict water activity coefficients at low acid mole fractions, while at high acid mole fractions both models fail to reproduce experimental activity coefficients. This indicates that COSMOtherm is not able to describe the water-acid interactions in supersaturated (crystalline) carboxylic acid mixtures.

Figure S4 of the Supplement shows comparisons between experimental and COSMOtherm-estimated water activity coefficients in oxalic, adipic and pimelic acid. For these three acids, only water activities have been determined experimentally (Braban et al., 2003; Marsh et al., 2017). In addition, COSMO-RS is likely sufficient in describing the acid-water interactions of oxalic acid, similar to the other dicarboxylic acids. Water activities in adipic and pimelic acid solutions were only measured in bulk solutions (Marsh et al., 2017). We found a good agreement between the bulk measurements and COSMO-RS-estimated water activity coefficients, with COSMOtherm slightly overestimating the experiments. This result is in line with previous comparisons of hydroxy carboxylic acids (Hytinen and Prisle, 2020).

The estimated COSMO-RS-estimated activity coefficients of the studied dicarboxylic acids are shown in Fig. 6. We can see that, using convention I, the activity coefficients of the smaller dicarboxylic acids are lower than for of the larger dicarboxylic acids. Water activity coefficients in all of the binary acid-water systems are close to 1 with  $x_{\text{acid}} < 0.1$ . For malonic ( $m = 3$ ) and glutaric ( $m = 5$ ) acid (with higher aqueous solubility), the activity coefficients deviate from ideality ( $\gamma = 1$ ) at higher acid mole fractions. Comparing COSMOtherm (solid lines) and UNIFAC estimates (dotted lines), there is less variation between the UNIFAC-estimated activity coefficients for the different acids studied than between the COSMOtherm estimates. This indicates that, in COSMOtherm, the number of carbon atoms has a larger effect on activity coefficients than estimated by UNIFAC.

We additionally calculated activity coefficients of oxalic acid (the most acidic dicarboxylic acid of this study) with dissociation of oxalic acid included in the system. In this case, the system contains neutral oxalic acid (HA) and water ( $\text{H}_2\text{O}$ ), as well as deprotonated oxalic acid ( $\text{A}^-$ ) and hydronium ion ( $\text{H}_3\text{O}^+$ ) according to the dissociation equilibrium



Figure S6 of the Supplement shows the difference between activity coefficients in a system where dissociation of oxalic acid is included and the binary system containing only neutral compounds. The calculation procedure is explained in more detail in the Supplement. There is no significant difference in the COSMO-RS and COSMO-RS-DARE activity coefficients of the dicarboxylic acid-water systems, the largest difference large difference in water activity coefficients when the ions are added to the system. A small change is seen in suberic and pimelic acids the acid activity coefficients, especially in the concentrated solutions where the estimated mole fraction of dissociated acid and hydronium ion is high. For the other carboxylic acids

studied here, the effect of including dissociation is likely to be smaller than for oxalic acid, due to the lower mole fractions of ions present in solutions of less acidic compounds.

### 3.3 Aqueous solubility

We estimated the aqueous solubility of the ~~carboxylic acids (monocarboxylic acids with  $n = 1-6$  and dicarboxylic acids with  $m = 5-8$ )~~ using the COSMO-RS-DARE method. Since activity coefficients are used in the equilibrium conditions of the LLE ~~and SLE~~ calculations, we used the same  $c_H$  parameters that were fitted in the activity coefficient calculations, to determine whether the same parameter value can be used in LLE ~~and SLE~~ calculations. As a comparison, we computed the same solubilities using COSMO-RS ~~and only conformers containing no intramolecular hydrogen bonds (dicarboxylic acids only)~~ (Kurtén et al., 2018; Hyttinen and Prisle, 2020). Based on previous COSMOtherm calculations, Michailoudi et al. (2020) found a good agreement with experimental aqueous solubilities of fatty acids with even number of carbon atoms ( $n = 2-12$ ). ~~In addition, Schröder et al. (2010) compared COSMOtherm estimated aqueous solubilities of dicarboxylic acids ( $m = 2-9$ ) with experiments and found that COSMOtherm is able to predict the temperature dependence quite well.~~ A comparison between experimentally determined aqueous solubilities and the COSMOtherm estimates of ~~mono- and dicarboxylic~~ monocarboxylic acids are shown in ~~Figs 7 and 8, respectively~~ Fig. 7.

We see that using COSMO-RS-DARE, COSMOtherm is able to predict the miscibility of the smaller monocarboxylic acids ( $n = 1-4$ ) but the experimental solubilities of pentanoic ( $n = 5$ ) and hexanoic ( $n = 6$ ) acids are overestimated to a greater degree than when using COSMO-RS. On the other hand, COSMO-RS underestimates the experimental solubility of butanoic acid by a factor of 18, while COSMO-RS-DARE overestimates the experimental solubilities (upper limit) of pentanoic and hexanoic acids only by factors of 3.4 and 4.1, respectively.

~~Different experimental enthalpy~~ For dicarboxylic acids, we estimated aqueous solubilities using COSMO-RS. The COSMOtherm-estimated and experimental solubilities are shown in Fig. 8. Different experimental heat of fusion and melting temperature point values have been reported for ~~the some of the studied~~ dicarboxylic acids. We calculated the lower and upper limit free energies of fusion by combining the different experimental values ~~and estimated the solubilities, and the aqueous solubilities were estimated~~ using the two different free energy of fusion values. ~~Using COSMO-RS, the~~ The higher  $\Delta G_{\text{fus}}$  estimate gives a lower aqueous solubility. ~~COSMOtherm overestimates the experimental aqueous solubility of oxalic acid ( $m = 2$ ) by a factor of 35 when the experimental melting point and heat of fusion by Booth et al. (2010) are used to estimate the free energy of fusion of oxalic acid. Cornils and Lappe (2000) and Omar and Ulrich (2006) also measured the melting point of oxalic acid and found temperatures almost 100 K higher than Booth et al. (2010). In addition, the heat of fusion measured by Omar and Ulrich (2006) is more than one order of magnitude higher than what was measured by Booth et al. (2010). This explains the differences in the estimated aqueous solubility of oxalic acid depending on whether the lower or upper limit estimate of the free energy of fusion is used. The variability in the COSMOtherm calculation. For the other acids, the variability in the measured melting points is much smaller (at most 11 K). For the SLE calculation using the COSMO-RS-DARE method, we used the upper limit  $\Delta G_{\text{fus}}$  estimate. -estimated solubilities is smaller than in the experimental solubilities.~~

Adding the acid dimerization in the system ( $m = 5-8$ ) has a smaller impact on the solubility estimate than the addition of the hydrated acid and water dimer ( $n = 4-6$ ). The COSMO-RS solubility estimates of most of the dicarboxylic acids ( $m = 3-7$ ) are within the ~~experimentally determined solubility range~~ range of experimentally determined solubilities. Using all ~~lowest-energy~~ lowest-energy conformers (up to 40 conformers), instead of only conformers containing no intramolecular hydrogen bonds, lowers the solubility estimates of all acids by a factor of 1.2 on average. The same effect of including conformers containing intramolecular hydrogen bonds has been previously seen in aqueous solubilities of citric, tartaric, malic and maleic acid, and multifunctional organosulfates (Hytinen and Prisle, 2020).

#### 4 Conclusions

We compared COSMO<sub>therm</sub>-estimated activity coefficients and aqueous solubilities of simple carboxylic acids with experimental values and a commonly used UNIFAC model, and found a good agreement between experiments and COSMO<sub>therm</sub> estimates. Using COSMO-RS-DARE, we were ~~able to significantly improve the activity coefficient estimates of water in~~ additionally able to improve the agreement between estimated and experimental water activity coefficient in binary monocarboxylic acid–water ~~systems~~ systems significantly compared to using COSMO-RS or UNIFAC. The COSMO-RS estimates of monocarboxylic acid activity coefficient in aqueous solutions agree with the experiments quite well, and ~~are~~ were further improved by COSMO-RS-DARE when the enthalpic fitting parameters were fitted using experimental activity coefficients. We were also able to estimate activity coefficients of pentanoic and hexanoic acids using only experimental water activity coefficients in the fitting of the COSMO-RS-DARE enthalpic parameters. In addition, COSMO-RS-DARE ~~is~~ was able to predict the miscibility of butanoic acid in water (using the fitting parameters of activity coefficient calculations), while COSMO-RS ~~predicts~~ predicted a finite solubility. However, in aqueous solubility calculations of pentanoic and hexanoic acid, COSMO-RS ~~leads~~ led to a better agreement between the experiments and COSMO<sub>therm</sub> estimates than COSMO-RS-DARE. For dicarboxylic acid–water systems, COSMO-RS produced better agreement with experiments than COSMO-RS-DARE. The experimental water activity coefficients from different sources have large variations and COSMO-RS-estimated water activity coefficients fit within the range of experimental water activity coefficients especially at low acid mole fractions ( $x_{\text{acid}} < 0.5$ ). We also found a good

425 agreement between COSMO-RS-estimated and experimental acid activity coefficients at all acid mole fractions.

COSMO<sub>therm</sub> ~~is~~ was able to reproduce the same even-odd behavior of the dicarboxylic acid properties that have previously been seen experimentally in vapor pressures (Bilde et al., 2003) and solubilities (Zhang et al., 2013), and computationally in gas-phase dimer formation (Elm et al., 2019). The calculated even-odd behavior observed here in aqueous solubilities is likely partially due to the even-odd variation of the melting points and heats of fusion. There is also no visible even-odd behavior in

430 the COSMO<sub>therm</sub>-estimated activity coefficients of the dicarboxylic acids. However, even-odd variation is seen in the effective equilibrium constants of dimerization of the larger dicarboxylic acids ( $m \geq 4$ ), which do not rely on experimental properties.

~~Cysewski (2019) found that COSMO-RS overestimates solubilities of ethenzamide in solvents such as ethyl acetate and methanol by up to a factor of 6 and a better estimates was found by fitting the enthalpic parameter in COSMO-RS-DARE to experiments. For the mono- and dicarboxylic acids, COSMO-RS over/underestimates the experimental solubility by up to a~~

435 factor of 3, with the exception of the solubility of butanoic ( $n = 4$ ), which is underestimated by a factor of 18. For each of the  
larger dicarboxylic acids ( $m = 5-8$ ), the solubility estimate changes by less than a factor of 1.6 between the COSMO-RS and  
COSMO-RS-DARE. Overall, the change in the COSMO-RS-DARE solubility compared to COSMO-RS solubility of mono-  
and dicarboxylic acids is not as large as was seen by Cysewski (2019), using the enthalpic parameter fitted to experimental  
activity coefficients.

440 Mono- and dicarboxylic acids are very common in the atmosphere and often used as model compounds for oxygenated  
functionalities in a range of applications from vapor pressure, condensation-evaporation, cloud condensation nuclei activity  
and hygroscopicity, but also aerosol phase and heterogeneous reactivity (Prenni et al., 2001; McNeill et al., 2008; Schwier  
et al., 2012; Rossignol et al., 2016). Solubilities and activity coefficients of these ~~SOA~~ secondary organic aerosol (SOA)  
constituents are needed to accurately predict their activities, to determine central properties such as composition, phase state,  
445 and chemical reactivity. Accurate computational tools are critical to provide this information for systems where experimental  
data are not readily accessible in literature or by experimental design. We showed that COSMOtherm provides a good solution  
to estimating thermodynamic properties of atmospherically relevant organic compounds that are not commercially available for  
measurements. In addition to simple binary systems studied here, COSMOtherm can be used to predict liquid-phase properties,  
such as activity coefficients, in complex, atmospherically relevant systems.

450 *Data availability.* The research data have been deposited in a reliable public data repository (the CERN Zenodo service) and can be accessed  
at <https://doi.org/10.5281/zenodo.3842593> (Hyttinen et al., 2020).

*Author contributions.* NH performed the COSMOtherm calculations. NH, RH and JE generated input files for the COSMOtherm calcula-  
tions. NH and NLP analyzed the results with contributions from all co-authors. NH wrote the first version of the manuscript and response to  
reviewers with contributions from all co-authors. NLP conceived, planned, supervised and secured funding for the project.

455 *Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* We thank CSC - IT Center for Science, Finland, for computational resources.

*Financial support.* This project has received funding from the European Research Council (ERC) under the European Union's Horizon  
2020 research and innovation programme, Project SURFACE (Grant Agreement No. 717022). Nønne L. Prisle, Noora Hyttinen, Reyhaneh  
Heshmatnezhad and Theo Kurtén also gratefully acknowledge the financial contribution from the Academy of Finland, including Grant

460 No. 308238, 314175 and 315600. Jonas Elm is grateful for financial support from the Swedish Research Council Formas project number 2018-01745-COBACCA.

## References

- AIOMFAC-web: version 2.32, <http://www.aiomfac.caltech.edu>, 2020.
- Aloisio, S., Hintze, P. E., and Vaida, V.: The hydration of formic acid, *J. Phys. Chem. A*, 106, 363–370, <https://doi.org/10.1021/jp012190l>,  
465 2002.
- Apelblat, A. and Manzurola, E.: Solubility of oxalic, malonic, succinic, adipic, maleic, malic, citric, and tartaric acids in water from 278.15 to 338.15 K, *J. Chem. Thermodyn.*, 19, 317–320, [https://doi.org/10.1016/0021-9614\(87\)90139-X](https://doi.org/10.1016/0021-9614(87)90139-X), 1987.
- Apelblat, A. and Manzurola, E.: Solubility of ascorbic, 2-furancarboxylic, glutaric, pimelic, salicylic, and o-phthalic acids in water from 279.15 to 342.15 K, and apparent molar volumes of ascorbic, glutaric, and pimelic acids in water at 298.15 K, *J. Chem. Thermodyn.*, 21,  
470 1005–1008, [https://doi.org/10.1016/0021-9614\(89\)90161-4](https://doi.org/10.1016/0021-9614(89)90161-4), 1989.
- Apelblat, A. and Manzurola, E.: Solubility of suberic, azelaic, levulinic, glycolic, and diglycolic acids in water from 278.25 K to 361.35 K, *J. Chem. Thermodyn.*, 22, 289–292, [https://doi.org/10.1016/0021-9614\(90\)90201-Z](https://doi.org/10.1016/0021-9614(90)90201-Z), 1990.
- Ben-Naim, A.: *Solvation Thermodynamics*, Plenum Press, New York and London, 1987.
- Bilde, M., Svenningsson, B., Mønster, J., and Rosenørn, T.: Even-odd alternation of evaporation rates and vapor pressures of C3–C9 dicarboxylic acid aerosols, *Environ. Sci. Technol.*, 37, 1371–1378, <https://doi.org/10.1021/es0201810>, 2003.  
475
- Booth, A. M., Barley, M. H., Topping, D. O., McFiggans, G., Garforth, A., and Percival, C. J.: Solid state and sub-cooled liquid vapour pressures of substituted dicarboxylic acids using Knudsen Effusion Mass Spectrometry (KEMS) and Differential Scanning Calorimetry, *Atmos. Chem. Phys.*, 10, 4879–4892, <https://doi.org/10.5194/acp-10-4879-2010>, 2010.
- Braban, C. F., Carroll, M. F., Styler, S. A., and Abbatt, J. P. D.: Phase transitions of malonic and oxalic acid aerosols, *J. Phys. Chem. A*, 107,  
480 6594–6602, <https://doi.org/10.1021/jp034483f>, 2003.
- Brooks, S. D., Wise, M. E., Cushing, M., and Tolbert, M. A.: Deliquescence behavior of organic/ammonium sulfate aerosol, *Geophys. Res. Lett.*, 29, 23–1, <https://doi.org/10.1029/2002GL014733>, 2002.
- Choi, M. Y. and Chan, C. K.: Continuous measurements of the water activities of aqueous droplets of water-soluble organic compounds, *J. Phys. Chem. A*, 106, 4566–4572, <https://doi.org/10.1021/jp013875o>, 2002.
- 485 Cornils, B. and Lappe, P.: Dicarboxylic acids, aliphatic, *Ullmann's Encyclopedia of Industrial Chemistry*, pp. 1–18, 2000.
- COSMObase: version 17.01, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2011.
- COSMOconf: version 4.3, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2013.
- COSMOtherm: version C3.0, Release 19, COSMOlogic GmbH & Co. KG., Leverkusen, Germany, 2019.
- Cysewski, P.: Prediction of ethenzamide solubility in organic solvents by explicit inclusions of intermolecular interactions within the framework of COSMO-RS-DARE, *J. Mol. Liq.*, p. 111163, <https://doi.org/10.1016/j.molliq.2019.111163>, 2019.  
490
- Davies, M. and Thomas, D. K.: Isopiestic Studies of Aqueous Dicarboxylic Acid Solutions, *J. Phys. Chem.*, 60, 41–44, <https://doi.org/10.1021/j150535a011>, 1956.
- Eckert, F. and Klamt, A.: Fast solvent screening via quantum chemistry: COSMO-RS approach, *AIChE J.*, 48, 369–385, <https://doi.org/10.1002/aic.690480220>, 2002.
- 495 Eckert, F. and Klamt, A.: *COSMOtherm Reference Manual*, version C30, Release 19, COSMOlogic GmbH & Co, KG.: Leverkusen, Germany, 2019.
- Elm, J., Kurtén, T., Bilde, M., and Mikkelsen, K. V.: Molecular interaction of pinic acid with sulfuric acid: Exploring the thermodynamic landscape of cluster growth, *J. Phys. Chem. A*, 118, 7892–7900, <https://doi.org/10.1021/jp503736s>, 2014.

- Elm, J., Hyttinen, N., Lin, J. J., Kurtén, T., and Prisle, N. L.: Strong Even/Odd Pattern in the Computed Gas-Phase Stability of Dicarboxylic Acid Dimers: Implications for Condensation Thermodynamics, *J. Phys. Chem. A*, 123, 9594–9599, <https://doi.org/10.1021/acs.jpca.9b08020>, 2019.
- Fisseha, R., Dommen, J., Gaeggeler, K., Weingartner, E., Samburova, V., Kalberer, M., and Baltensperger, U.: Online gas and aerosol measurement of water soluble carboxylic acids in Zurich, *J. Geophys. Res.: Atmos.*, 111, D12 316, <https://doi.org/10.1029/2005JD006782>, 2006.
- Fredenslund, A., Jones, R. L., and Prausnitz, J. M.: Group-contribution estimation of activity coefficients in nonideal liquid mixtures, *AIChE J.*, 21, 1086–1099, <https://doi.org/10.1002/aic.690210607>, 1975.
- Guo, H., Zhou, J., Wang, L., Zhou, Y., Yuan, J., Zhao, R., et al.: Seasonal variations and sources of carboxylic acids in PM<sub>2.5</sub> in Wuhan, China, *Aerosol Air Qual. Res.*, 15, 517–528, <https://doi.org/10.4209/aaqr.2014.02.0040>, 2014.
- Hansen, R. S., Miller, F. A., and Christian, S. D.: Activity Coefficients of Components in the Systems Water–Acetic Acid, Water–Propionic Acid and Water–*n*-Butyric Acid at 25°, *J. Phys. Chem.*, 59, 391–395, <https://doi.org/10.1021/j150527a003>, 1955.
- Hyder, M., Genberg, J., Sandahl, M., Swietlicki, E., and Jönsson, J. Å.: Yearly trend of dicarboxylic acids in organic aerosols from south of Sweden and source attribution, *Atmos. Environ.*, 57, 197–204, <https://doi.org/10.1016/j.atmosenv.2012.04.027>, 2012.
- Hyttinen, N. and Prisle, N. L.: Improving solubility and activity estimates of multifunctional atmospheric organics by selecting conformers in COSMOtherm, *J. Phys. Chem. A*, 124, 4801–4812, <https://doi.org/10.1021/acs.jpca.0c04285>, 2020.
- Hyttinen, N., Heshmatnezhad, R., Elm, J., Kurtén, T., and Prisle, N. L.: Supplementary data for the manuscript "Technical note: Estimating aqueous solubilities and activity coefficients of mono- and  $\alpha,\omega$ -dicarboxylic acids using COSMO-RS-DARE" (Version 1) [Data set], Zenodo, <https://doi.org/10.5281/zenodo.3842593>, 2020.
- Jones, E. R. and Bury, C. R.: LXXVII. The freezing-points of concentrated solutions.–Part II. Solutions of formic, acetic, propionic, and butyric acids, *London, Edinburgh Dublin Philos. Mag. J. Sci.*, 4, 841–848, <https://doi.org/10.1080/14786441008564388>, 1927.
- Jung, J., Tsatsral, B., Kim, Y. J., and Kawamura, K.: Organic and inorganic aerosol compositions in Ulaanbaatar, Mongolia, during the cold winter of 2007 to 2008: dicarboxylic acids, ketocarboxylic acids, and  $\alpha$ -dicarbonyls, *J. Geophys. Res.: Atmos.*, 115, D22 203, <https://doi.org/10.1029/2010JD014339>, 2010.
- Kawamura, K. and Sakaguchi, F.: Molecular distributions of water soluble dicarboxylic acids in marine aerosols over the Pacific Ocean including tropics, *J. Geophys. Res.: Atmos.*, 104, 3501–3509, <https://doi.org/10.1029/1998JD100041>, 1999.
- Kawamura, K., Seméré, R., Imai, Y., Fujii, Y., and Hayashi, M.: Water soluble dicarboxylic acids and related compounds in Antarctic aerosols, *J. Geophys. Res.: Atmos.*, 101, 18 721–18 728, <https://doi.org/10.1029/96JD01541>, 1996.
- Kawamura, K., Steinberg, S., and Kaplan, I. R.: Homologous series of C1–C10 monocarboxylic acids and C1–C6 carbonyls in Los Angeles air and motor vehicle exhausts, *Atmos. Environ.*, 34, 4175–4191, [https://doi.org/10.1016/S1352-2310\(00\)00212-0](https://doi.org/10.1016/S1352-2310(00)00212-0), 2000.
- Kildgaard, J. V., Mikkelsen, K. V., Bilde, M., and Elm, J.: Hydration of Atmospheric Molecular Clusters II: Organic Acid–Water Clusters, *J. Phys. Chem. A*, 122, 8549–8556, <https://doi.org/10.1021/acs.jpca.8b07713>, 2018.
- Klamt, A.: Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.*, 99, 2224–2235, <https://doi.org/10.1021/j100007a062>, 1995.
- Klamt, A. and Schüürmann, G.: COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, *J. Chem. Soc., Perkin Trans. 2*, pp. 799–805, <https://doi.org/10.1039/P29930000799>, 1993.
- Klamt, A., Jonas, V., Bürger, T., and Lohrenz, J. C. W.: Refinement and parametrization of COSMO-RS, *J. Phys. Chem. A*, 102, 5074–5085, <https://doi.org/10.1021/jp980017s>, 1998.



- Kurtén, T., Hyttinen, N., D'Ambro, E. L., Thornton, J., and Prisle, N. L.: Estimating the saturation vapor pressures of isoprene oxidation products  $C_5H_{12}O_6$  and  $C_5H_{10}O_6$  using COSMO-RS, *Atmos. Chem. Phys.*, 18, 17 589–17 600, <https://doi.org/10.5194/acp-18-17589-2018>, 2018.
- 540 Levine, I. N.: *Physical Chemistry*, McGraw-Hill, New York, USA, 6th edn., 2009.
- Maffia, M. C. and Meirelles, A. J. A.: Water activity and pH in aqueous polycarboxylic acid systems, *J. Chem. Eng. Data*, 46, 582–587, <https://doi.org/10.1021/je0002890>, 2001.
- Marsh, A., Miles, R. E. H., Rovelli, G., Cowling, A. G., Nandy, L., Dutcher, C. S., and Reid, J. P.: Influence of organic compound functionality on aerosol hygroscopicity: dicarboxylic acids, alkyl-substituents, sugars and amino acids, *Atmos. Chem. Phys.*, 17, 5583, <https://doi.org/10.5194/acp-17-5583-2017>, 2017.
- 545 McNeill, V. F., Yatavelli, R. L. N., Thornton, J. A., Stipe, C. B., and Landgrebe, O.: Heterogeneous OH oxidation of palmitic acid in single component and internally mixed aerosol particles: vaporization and the role of particle phase, *Atmos. Chem. Phys.*, 8, 5465–5478, <https://doi.org/10.5194/acp-8-5465-2008>, 2008.
- Michailoudi, G., Hyttinen, N., Kurtén, T., and Prisle, N. L.: Solubility and Activity Coefficients of Atmospheric Surfactants in Aqueous Solution Evaluated using COSMOtherm, *J. Phys. Chem. A*, 124, 430–443, <https://doi.org/10.1021/acs.jpca.9b09780>, 2020.
- Mochida, M., Kawamura, K., Umemoto, N., Kobayashi, M., Matsunaga, S., Lim, H.-J., Turpin, B. J., Bates, T. S., and Simoneit, B. R.: Spatial distributions of oxygenated organic compounds (dicarboxylic acids, fatty acids, and levoglucosan) in marine aerosols over the western Pacific and off the coast of East Asia: Continental outflow of organic aerosols during the ACE-Asia campaign, *J. Geophys. Res.: Atmos.*, 108, 8638, <https://doi.org/10.1029/2002JD003249>, 2003.
- 555 Omar, W. and Ulrich, J.: Solid liquid equilibrium, metastable zone, and nucleation parameters of the oxalic acid- water system, *Cryst. Growth Des.*, 6, 1927–1930, <https://doi.org/10.1021/cg060112n>, 2006.
- O'Neil, M. J.: *The Merck index: an encyclopedia of chemicals, drugs, and biologicals*, RSC Publishing, 2013.
- Peng, C., Chan, M. N., and Chan, C. K.: The hygroscopic properties of dicarboxylic and multifunctional acids: Measurements and UNIFAC predictions, *Environ. Sci. Technol.*, 35, 4495–4501, <https://doi.org/10.1021/es0107531>, 2001.
- 560 Prenni, A. J., DeMott, P. J., Kreidenweis, S. M., Sherman, D. E., Russell, L. M., and Ming, Y.: The effects of low molecular weight dicarboxylic acids on cloud formation, *J. Phys. Chem. A*, 105, 11 240–11 248, <https://doi.org/10.1021/jp012427d>, 2001.
- Prigogine, I. and Defay, R.: *Chemical Thermodynamics*, Longmans Green, 1st edition edn., 1954.
- Pye, H. O. T., Nenes, A., Alexander, B., Ault, A. P., Barth, M. C., Clegg, S. L., Collett Jr., J. L., Fahey, K. M., Hennigan, C. J., Herrmann, H., Kanakidou, M., Kelly, J. T., Ku, I.-T., McNeill, V. F., Riemer, N., Schaefer, T., Shi, G., Tilgner, A., Walker, J. T., Wang, T., Weber, R., Xing, J., Zaveri, R. A., and Zuend, A.: The acidity of atmospheric particles and clouds, *Atmos. Chem. Phys.*, 20, 4809–4888, <https://doi.org/10.5194/acp-20-4809-2020>, 2020.
- 565 Romero, C. M. and Suárez, F.: Effect of temperature on the solubility of short-chain carboxylic acids in water, *J. Solution Chem.*, 38, 315–320, <https://doi.org/10.1007/s10953-009-9375-6>, 2009.
- Rossignol, S., Tinel, L., Bianco, A., Passananti, M., Brigante, M., Donaldson, D. J., and George, C.: Atmospheric photochemistry at a fatty acid-coated air-water interface, *Science*, 353, 699–702, <https://doi.org/10.1126/science.aaf3617>, 2016.
- 570 Roux, M. V., Temprado, M., and Chickos, J. S.: Vaporization, fusion and sublimation enthalpies of the dicarboxylic acids from  $C_4$  to  $C_{14}$  and  $C_{16}$ , *J. Chem. Thermodyn.*, 37, 941–953, <https://doi.org/10.1016/j.jct.2004.12.011>, 2005.
- Sachsenhauser, T., Rehfeldt, S., Klamt, A., Eckert, F., and Klein, H.: Consideration of dimerization for property prediction with COSMO-RS-DARE, *Fluid Phase Equilib.*, 382, 89–99, <https://doi.org/10.1016/j.fluid.2014.08.030>, 2014.

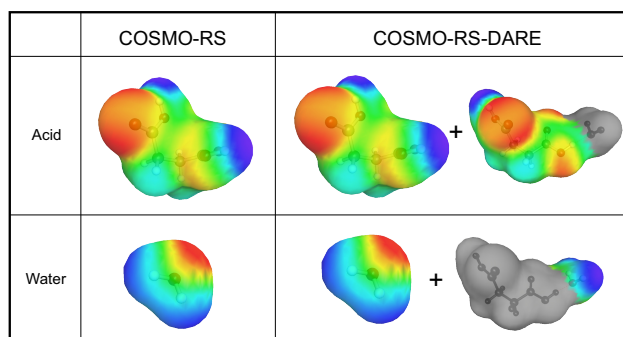
- 575 Saxena, P. and Hildemann, L. M.: Water-soluble organics in atmospheric particles: A critical review of the literature and application of thermodynamics to identify candidate compounds, *J. Atmos. Chem.*, 24, 57–109, <https://doi.org/10.1007/BF00053823>, 1996.
- Schröder, B., Santos, L. M. N. B. F., Marrucho, I. M., and Coutinho, J. A. P.: Prediction of aqueous solubilities of solid carboxylic acids with COSMO-RS, *Fluid Phase Equilib.*, 289, 140–147, <https://doi.org/10.1016/j.fluid.2009.11.018>, 2010.
- Schwier, A., Mitroo, D., and McNeill, V. F.: Surface tension depression by low-solubility organic material in aqueous aerosol mimics, *Atmos. Environ.*, 54, 490–495, <https://doi.org/10.1016/j.atmosenv.2012.02.032>, 2012.
- 580 Song, M., Marcolli, C., Krieger, U. K., Zuend, A., and Peter, T.: Liquid-liquid phase separation and morphology of internally mixed dicarboxylic acids/ammonium sulfate/water particles, *Atmos. Chem. Phys.*, 12, 2691–2712, <https://doi.org/10.5194/acp-12-2691-2012>, 2012.
- Soonsin, V., Zardini, A. A., Marcolli, C., Zuend, A., and Krieger, U. K.: The vapor pressures and activities of dicarboxylic acids reconsidered: the impact of the physical state of the aerosol, *Atmos. Chem. Phys.*, 10, 11 753–11 767, <https://doi.org/10.5194/acp-10-11753-2010>, 2010.
- 585 Sordo, J. Á.: Solvation thermodynamics: two formulations and some misunderstandings, *RSC Adv.*, 5, 96 105–96 116, <https://doi.org/10.1039/c5ra17305a>, 2015.
- Tsai, Y. I. and Kuo, S.-C.: Contributions of low molecular weight carboxylic acids to aerosols and wet deposition in a natural subtropical broad-leaved forest environment, *Atmos. Environ.*, 81, 270–279, <https://doi.org/10.1016/j.atmosenv.2013.08.061>, 2013.
- TURBOMOLE: Version 7.11, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, TURBOMOLE GmbH, 590 2010.
- Vawdrey, A. C., Oscarson, J. L., Rowley, R. L., and Wilding, W. V.: Vapor-phase association of n-aliphatic carboxylic acids, *Fluid Phase Equilib.*, 222, 239–245, <https://doi.org/10.1016/j.fluid.2004.06.043>, 2004.
- Verma, N., Satsangi, A., Lakhani, A., and Kumari, K. M.: Low molecular weight monocarboxylic acids in PM<sub>2.5</sub> and PM<sub>10</sub>: Quantification, seasonal variation and source apportionment, *Aerosol Air Qual. Res.*, 17, 485–498, <https://doi.org/10.4209/aaqr.2016.05.0183>, 2017.
- 595 Wavefunction Inc.: Spartan'14, Irvine, CA, 2014.
- Wavefunction Inc.: Spartan'16, Irvine, CA, 2016.
- Weber, K. H., Morales, F. J., and Tao, F.-M.: Theoretical study on the structure and stabilities of molecular clusters of oxalic acid with water, *J. Phys. Chem. A*, 116, 11 601–11 617, <https://doi.org/10.1021/jp308499f>, 2012.
- Weber, R. J., Guo, H., Russell, A. G., and Nenes, A.: High aerosol acidity despite declining atmospheric sulfate concentrations over the past 600 15 years, *Nat. Geosci.*, 9, 282–285, <https://doi.org/10.1038/ngeo2665>, 2016.
- Wise, M. E., Surratt, J. D., Curtis, D. B., Shilling, J. E., and Tolbert, M. A.: Hygroscopic growth of ammonium sulfate/dicarboxylic acids, *J. Geophys. Res.: Atmos.*, 108, 4638–4642, <https://doi.org/10.1029/2003JD003775>, 2003.
- Zhang, H., Xie, C., Liu, Z., Gong, J., Bao, Y., Zhang, M., Hao, H., Hou, B., and Yin, Q.-x.: Identification and molecular understanding of the odd–even effect of dicarboxylic acids aqueous solubility, *Ind. Eng. Chem. Res.*, 52, 18 458–18 465, <https://doi.org/10.1021/ie4030837>, 605 2013.
- Zhao, W., Kawamura, K., Yue, S., Wei, L., Ren, H., Yan, Y., Kang, M., Li, L., Ren, L., Lai, S., Li, Y., Wang, Z., and Fu, P.: Molecular distribution and compound-specific stable carbon isotopic composition of dicarboxylic acids, oxocarboxylic acids and  $\alpha$ -dicarbonyls in PM<sub>2.5</sub> from Beijing, China, *Atmos. Chem. Phys.*, 18, 2749–2767, <https://doi.org/10.5194/acp-18-2749-2018>, 2018.
- Zuend, A., Marcolli, C., Luo, B. P., and Peter, T.: A thermodynamic model of mixed organic-inorganic aerosols to predict activity coefficients, 610 *Atmos. Chem. Phys.*, 8, 4559–4593, <https://doi.org/10.5194/acp-8-4559-2008>, 2008.
- Zuend, A., Marcolli, C., Booth, A. M., Lienhard, D. M., Soonsin, V., Krieger, U. K., Topping, D. O., McFiggans, G., Peter, T., and Seinfeld, J. H.: New and extended parameterization of the thermodynamic model AIOMFAC: calculation of activity coefficients for organic-

inorganic mixtures containing carboxyl, hydroxyl, carbonyl, ether, ester, alkenyl, alkyl, and aromatic functional groups, *Atmos. Chem. Phys.*, 11, 9155–9206, <https://doi.org/10.5194/acp-11-9155-2011>, 2011.

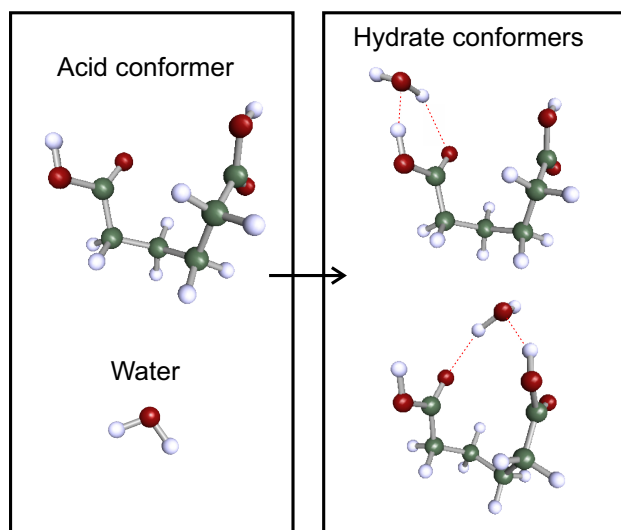
**Table 1.** List of the studied  $\alpha,\omega$ -dicarboxylic acids and their ~~literature values of experimentally determined~~ melting ~~point points~~ and ~~heat~~ heats of fusion. ~~The values were measured using anhydrous acids (crystalline at 298.15 K). The melting point of oxalic acid reported by (Booth et al., 2010) (370 K) is likely the temperature of the phase transition from dihydrate to anhydrous crystal polymorph. Similar transition was seen by (Omar and Ulrich, 2006) at 378.35 K in their differential scanning calorimetry experiment.~~

| Acid     | Chemical formula                              | Melting point (K)  | Heat of fusion (kJ mol <sup>-1</sup> )              |
|----------|---|--|---|
| Oxalic   | C <sub>2</sub> H <sub>2</sub> O <sub>4</sub>  | <del>370<sup>a</sup></del> <u>462.65<sup>a</sup></u> ; 465.26 <sup>b</sup> ; <del>462.65<sup>c</sup></del> | <del>3.424<sup>a</sup></del> ; 58.158 <sup>b</sup>  |
| Malonic  | C <sub>3</sub> H <sub>4</sub> O <sub>4</sub>  | 406 <sup>a,c</sup> ; 408.15 <sup>c,a</sup>   | 18.739 <sup>a,c</sup>                               |
| Succinic | C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>  | <u>455.2<sup>d</sup></u> ; <u>458<sup>a,c</sup></u> ; <u>461.15<sup>c,a</sup></u>                          | 31.259 <sup>a,c</sup> ; <u>34.0<sup>d</sup></u>     |
| Glutaric | C <sub>5</sub> H <sub>8</sub> O <sub>4</sub>  | <u>363.9<sup>d</sup></u> ; <u>369<sup>a,c</sup></u> ; <u>372.15<sup>c,a</sup></u>                          | <u>18.8<sup>d</sup></u> ; 22.043 <sup>a,c</sup>     |
| Adipic   | C <sub>6</sub> H <sub>10</sub> O <sub>4</sub> | <u>419.0<sup>d</sup></u> ; <u>423<sup>a,c</sup></u> ; <u>426.15<sup>c,a</sup></u>                          | <u>33.7<sup>d</sup></u> ; 35.891 <sup>a,c</sup>     |
| Pimelic  | C <sub>7</sub> H <sub>12</sub> O <sub>4</sub> | <del>368</del> <u>368.2<sup>d</sup></u> ; 379.15 <sup>c,a</sup>  | <del>28.9<sup>d</sup></del> <u>23.7<sup>d</sup></u> |
| Suberic  | C <sub>8</sub> H <sub>14</sub> O <sub>4</sub> | <del>413</del> <u>413.2<sup>d</sup></u> ; 417.15 <sup>c,a</sup>  | <del>38.2</del> <u>30.7<sup>d</sup></u>             |

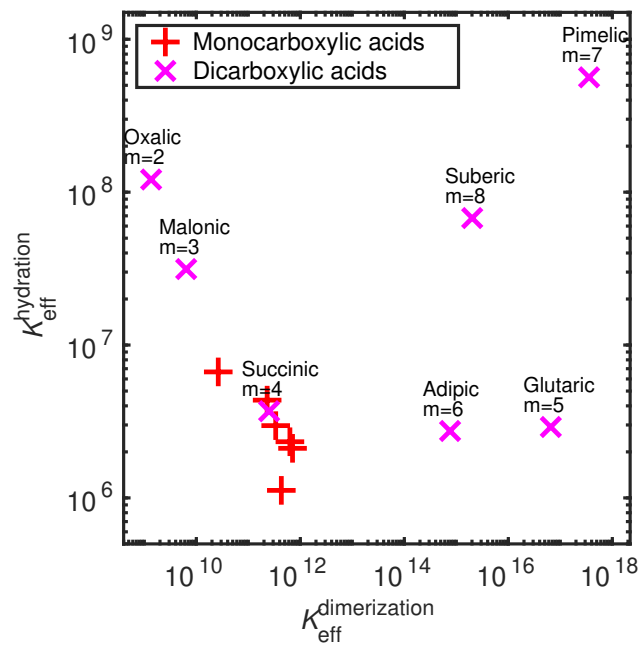
<sup>a</sup>~~Booth et al. (2010)~~ Cornils and Lappe (2000); <sup>b</sup>Omar and Ulrich (2006); <sup>c</sup>~~Cornils and Lappe (2000)~~ Booth et al. (2010); <sup>d</sup>Roux et al. (2005)



**Figure 1.** The  $\sigma$ -surfaces of succinic acid and water conformers used in COSMO-RS and COSMO-RS-DARE calculations. The conformer distributions in COSMO-RS-DARE include parts of cluster  $\sigma$ -surfaces (in this example a hydrate cluster). Color coding of  $\sigma$ -surfaces: red = negative partial charge, blue = positive partial charge, green = neutral partial charge, grey = omitted  $\sigma$ -surface.



**Figure 2.** [The formation of dicarboxylic acid hydrate conformers. Color coding: green = C, white = H, red = O.](#)

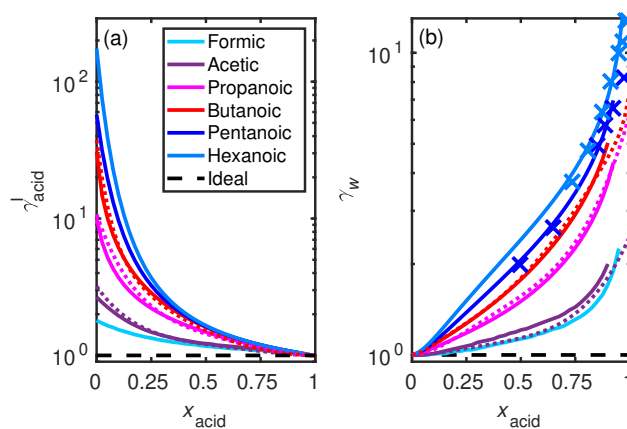


**Figure 3.** Dimensionless effective equilibrium constants of acid dimer and hydrate formation ( $K_{\text{eff}}^{\text{dimerization}}$  and  $K_{\text{eff}}^{\text{hydration}}$ , respectively, in mole-fraction-basis) in condensed phase, at 298.15 K. See Table 2 for the values.

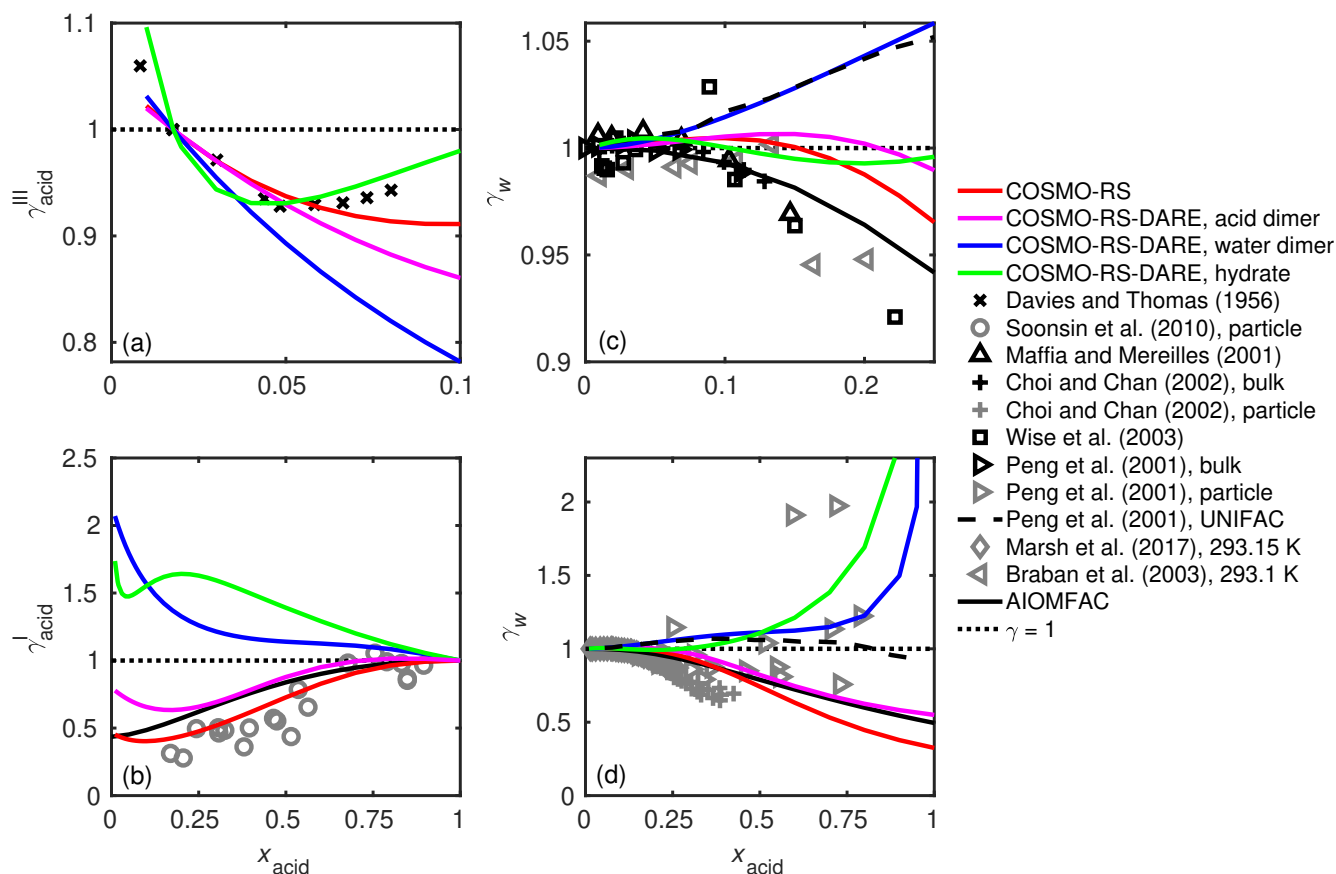
**Table 2.** Dimensionless effective equilibrium constants ( $K_{\text{eff}}$  ~~in mole fraction basis~~) of cluster formation in condensed phase, at 298.15 K.

| Acid      | Carbon # | $K_{\text{eff}}^{\text{hydration}}$ | $K_{\text{eff}}^{\text{dimerization}}$ |
|-----------|----------|-------------------------------------|--|
| Formic    | 1        | $6.67 \times 10^6$                  | $2.64 \times 10^{10}$                  |
| Acetic    | 2        | $4.36 \times 10^6$                  | $2.30 \times 10^{11}$                  |
| Propanoic | 3        | $1.12 \times 10^6$                  | $4.31 \times 10^{11}$                  |
| Butanoic  | 4        | $2.97 \times 10^6$                  | $3.36 \times 10^{11}$                  |
| Pentanoic | 5        | $2.11 \times 10^6$                  | $7.08 \times 10^{11}$                  |
| Hexanoic  | 6        | $2.33 \times 10^6$                  | $6.30 \times 10^{11}$                  |
| Oxalic    | 2        | $1.21 \times 10^8$                  | $1.35 \times 10^9$                     |
| Malonic   | 3        | $3.14 \times 10^7$                  | $6.38 \times 10^9$                     |
| Succinic  | 4        | $3.69 \times 10^6$                  | $2.51 \times 10^{11}$                  |
| Glutaric  | 5        | $2.91 \times 10^6$                  | $6.56 \times 10^{16}$                  |
| Adipic    | 6        | $2.74 \times 10^6$                  | $7.58 \times 10^{14}$                  |
| Pimelic   | 7        | $5.64 \times 10^8$                  | $3.57 \times 10^{17}$                  |
| Suberic   | 8        | $6.77 \times 10^7$                  | $2.02 \times 10^{15}$                  |
| Water     | -        | -                                   | $5.71 \times 10^5$                     |

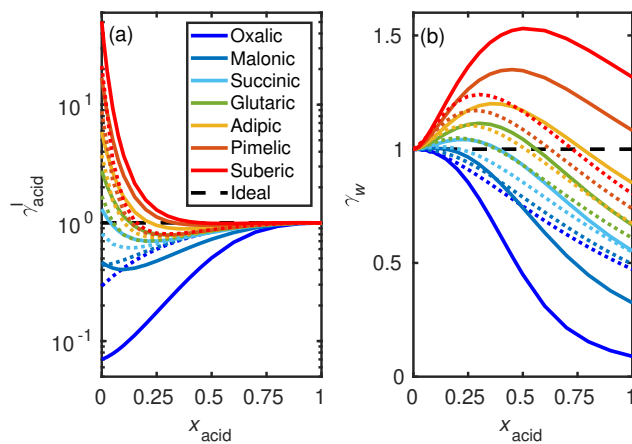




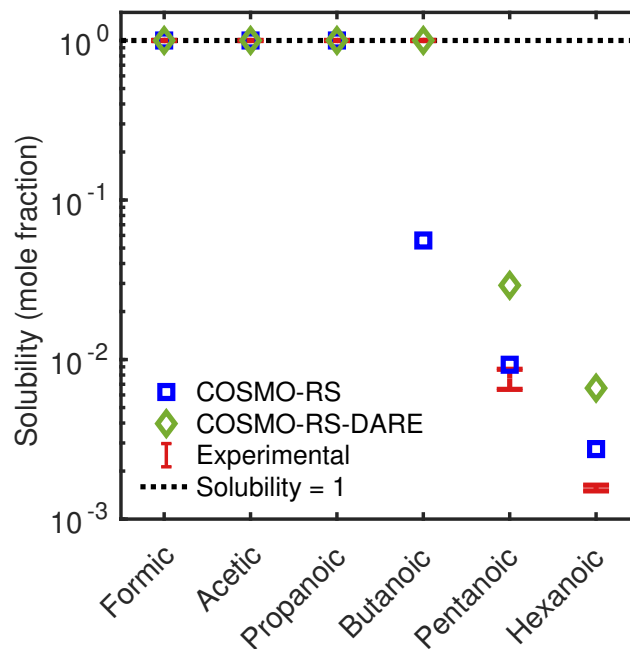
**Figure 4.** Activity coefficients ( $\gamma^I$ ) of the monocarboxylic acids (left panel) monocarboxylic acids and water (right panel) water in all mixing states of the binary aqueous solutions, at 298.15 K. The solid lines represent activity coefficient estimates using COSMO-RS-DARE, dotted lines are calculated from the equations fitted to experiments by Hansen et al. (1955), and the markers are the experimental points from the same study. For the studied acids with finite aqueous solubilities at 298.15 K (pentanoic and hexanoic acid), water activity coefficients were measured using acid-rich solutions ((Hansen et al., 1955)). The water activity coefficients at high  $x_{\text{acid}}$  are not shown in the figure, because COSMO-RS-DARE overpredicts the experiments by several orders of magnitude. All activity coefficient values are given in Tables S3 and S4 of the Supplement.



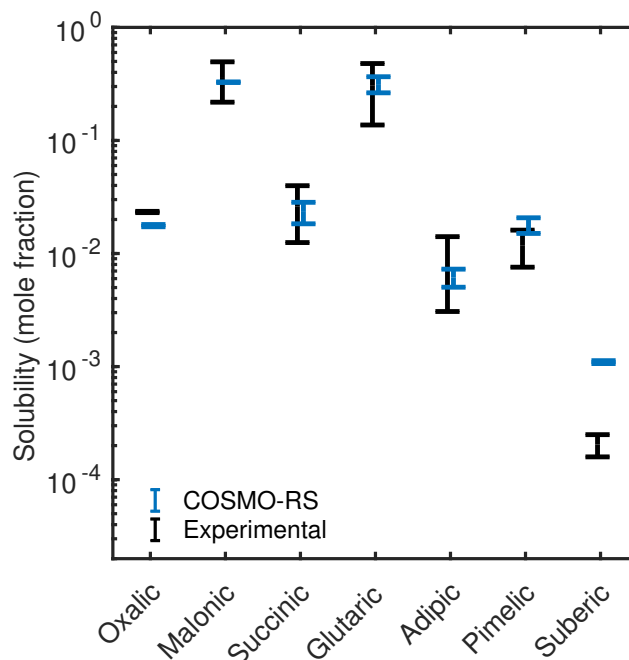
**Figure 5.** *COSMOtherm* estimated activity coefficients ( $\gamma^{\text{I}}$ ) of dicarboxylic acids (left panels a-b) malonic acid and water (right panels c-d) water in the binary acid-water mixtures at 298.15 K. Solid lines represent COSMO-RS estimates and dotted lines COSMO-RS-DARE estimates with the acid dimers included calculated using different clustering reactions in the COSMO-RS-DARE calculation. The activity coefficients were estimated up to the experimental solubility limit of the acids malonic acid by Davies and Thomas (1956) ( $x_{\text{acid}} \leq x_{\text{SOL,acid}}$  at 298.15 K given in convention III) and Soonsin et al. (2010) (particle measurements at various temperatures given in Tables S5-convention I) and S6 of the Supplement by Maffia and Meirelles (2001), Choi and Chan (2002), Wise et al. (2003), Peng et al. (2001), Marsh et al. (2017), Braban et al. (2003) and AIOMFAC-web (2020).



**Figure 6.** COSMO-RS- (solid lines) and UNIFAC-estimated (dotted lines; AIOMFAC-web, 2020) activity coefficients ( $\gamma^I$ ) of (a) dicarboxylic acids and (b) water in the binary acid–water mixtures at 298.15 K. All COSMO $therm$ -estimated activity coefficient values are given in Tables S5 and S6 of the Supplement.



**Figure 7.** COSMO~~therm~~~~estimated~~-~~estimated~~ aqueous solubility of monocarboxylic acids calculated using COSMO-RS and COSMO-RS-DARE, at 298.15 K. Experimental solubilities by Saxena and Hildemann (1996) ( $n = 1-6$ ) and Romero and Suárez (2009) ( $n = 5-6$ ).



**Figure 8.** Estimated solubilities of the dicarboxylic acids compared to their experimental solubilities by [Apelblat and Manzurola \(1987\)](#) (oxalic — [\(Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Omar and Ulrich, 2006; Zhang et al., 2013\)](#)  $m = 2-4$ , malonic — [\(Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Zhang et al., 2013\)](#)6), succinic [\(Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Brooks et al., 2002; Zhang et al., 2013\)](#) [Apelblat and Manzurola \(1989\)](#) ( $m = 5$ ), glutaric — [\(Apelblat and Manzurola, 1989; Saxena and Hildemann, 1996; Song et al., 2012; O’Neil, 2013; Zhang et al., 2013\)](#) [Apelblat and Manzurola \(1990\)](#) ( $m = 7-8$ ), adipic — [\(Apelblat and Manzurola, 1987; Saxena and Hildemann, 1996; Zhang et al., 2013\)](#) [Saxena and Hildemann \(1996\)](#) ( $m = 2-6$ ), pimelic — [\(Apelblat and Manzurola, 1990; Zhang et al., 2013\)](#) [Omar and Ulrich \(2006\)](#) ( $m = 2$ ), [Zhang et al. \(2013\)](#) ( $m = 2-8$ ), [Brooks et al. \(2002\)](#) ( $m = 4$ ), [Song et al. \(2012\)](#) ( $m = 5$ ) and suberic [\(Apelblat and Manzurola, 1990; Zhang et al., 2013\)](#) [O’Neil \(2013\)](#) ( $m = 5$ ), at 298.15 K. The lower and upper limit solubility estimates are obtained using the highest and lowest free energies of fusion (estimated from experimental melting point and heat of fusion in Table 1), respectively.