**Atmospheric
Chemistry
and Physics
Discussions**

Interactive
Comment

Interactive comment on "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" *by* A. Zuend et al.

A. Zuend et al.

zuend@caltech.edu

Received and published: 8 August 2011

The authors would like to thank referee #1 for the careful review of the manuscript and valuable comments. We have addressed the reviewer's general and specific comments below. We will also use the points raised to clarify our statements in the revised manuscript version.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

General comments

"My only important remark is on the use of the existing UNIFAC parameters, assembled from different research groups, for describing the organic-organic and organic-water interactions. The authors have stressed the importance of fitting the organic-ion interaction parameters for all data simultaneously. Shouldn't the same be true for the organic-organic and organic-water interactions, i.e. wouldn't it be better to refit these parameters over a wide variety of mixtures at once? The authors point to the shortcomings of the UNIFAC model in some circumstances, even for salt-free systems, and blame this solely on the group-contribution concept. But I think there can be another reason: the fact that the UNIFAC set of interaction parameters is not obtained by an overall optimisation. It would make sense to me -if time and manpower would be no limitation- to first have the optimal UNIFAC parameters for a best description of salt-free mixtures, before improving/extending the organic-ion interaction parameters. Have the authors refrained from this because of the challenge of the task, or do they think that the existing UNIFAC parameters cannot be much improved upon?"

The original UNIFAC model version of Fredenslund et al. (1975) and subsequent revisions and improvements, e.g., by Hansen et al. (1991) (standard UNIFAC), have been fitted to a very large database of organic-organic and organic-water data (mostly VLE data). Revisions and new UNIFAC parameters are compatible with the original UNIFAC parameter set as long as the same mathematical model expressions are used. This is the reason why some new UNIFAC parameter sets can be used in combination with the original model, while other parameter sets, e.g., those for UNIFAC Dortmund are not compatible. One of the shortcomings of the standard UNIFAC in terms of the group-contribution approach and the data used for its simultaneous parameter fit, is the large bias of the used database towards small, monofunctional compounds and temperatures at and mainly above room temperature. However, as by now, most

thermodynamic data available in the literature are still for organic compounds with one or two functional groups. As described in Section 3.1 of the article, in AIOMFAC the additional/revised UNIFAC parameters include the improved sets of Marcolli and Peter (2005) (UNIFAC-MP) and of Peng et al. (2001) (UNIFAC-Peng) for alcohols/polyols and carboxylic/dicarboxylic acids, respectively. These parameter sets are compatible with the original UNIFAC and differ only in a few interaction parameters that have been improved with new experimental data in the respective studies, while at the same time the other parameters were kept as given by Hansen et al. (1991). We tested different options of combining the original parameters together with the revised UNIFAC parameters within AIOMFAC. We found that combining original UNIFAC with parameters of Peng et al. (2001) for $COOH$ interactions with water and OH groups and the parameters of Marcolli and Peter (2005) for interactions of OH groups with water (plus their introduction of specific CH_n groups and parameters for alcohols) lead in most cases to similar or better results than when using standard UNIFAC only. For example, the combination of interaction parameters of standard UNIFAC (Hansen et al.), UNIFAC-Peng, and UNIFAC-MP in solutions containing water and citric acid (see Fig. 1 of the article) leads only to small deviations as compared to using only UNIFAC-Peng (for both $COOH$ and OH) and standard UNIFAC. Tests also showed that the combination of the three sources of UNIFAC parameters seems not to be the reason for large deviations from experimental data found for some mixtures. Rather, this is in some cases really a limitation of the group-contribution concept (that could, in principle, be overcome by introducing additional parameters, e.g., accounting for proximity effects of functional groups). Our estimate is, that for many mixtures used in the study, a new simultaneous fit of all UNIFAC parameters would likely only lead to little improvement of UNIFAC at room temperature and higher temperatures. However, the UNIFAC performance for atmospherically relevant species to lower temperatures can be improved with additional data and a new fit. This is work in progress.

"The complexity of UNIFAC/AIOMFAC stands in sharp contrast with the simple one of

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Donahue et al. (2011). This model only aims to model the overall aerosol properties, not the individual components. Can this simple model be sufficient to predict e.g. phase separation in aerosol, or do the authors think that a detailed model such as UNIFAC/AIOMFAC is indispensable for this?"

The model of Donahue et al. (2011) uses a very different approach with expressions of a much simpler form to account for overall non-ideality. In contrast to a detailed group-contribution model like AIOMFAC, the model of Donahue et al. (2011) is only meant to predict average activity properties of organic mixtures (pseudo-binary approach) based on knowledge of average O:C and H:C ratios, rather than individual activity coefficients of all compounds. Hence, the two models have a quite different level of complexity in terms of how they represent the thermodynamics of real mixtures. The two models have also different practical applications in terms of their requirements for and usage of given system information. One of the general differences between the Donahue et al. model and AIOMFAC is that AIOMFAC uses/requires functional group information of organics as input and includes water and dissolved inorganic ions in the framework, while the model of Donahue et al. (2011) is for mixtures of organics only. Since one important way by which a liquid-liquid phase separation is triggered is through the interactions of ions, water, and organics, especially by means of salting-out of the organics, we believe that a somewhat accurate prediction of phase separation requires consideration of these interactions (in mixed aerosol particles). Whether the model of Donahue et al. (2011) can predict liquid-liquid phase separation (in organic mixtures only) with any reasonable accuracy is not shown in their article. Their model has not been tested or parameterized with measured activity data; AIOMFAC on the other hand is systematically fitted and compared to an extensive database of experimental data. Concerning atmospheric aerosols, different applications of the two models can also be described this way: for an aerosol that is only characterized by its volatility and O:C, the model of Donahue et al. can be used to roughly estimate/constrain the numbers and ratios of carbon, hydrogen, and oxygen

atoms of an average compound in the (organic) mixture; if more is known about the composition or if a reasonable estimation of a set of representative compounds can be made, AIOMFAC can be used to quite accurately calculate compound-specific activity coefficients and to predict a potential phase separation. For the calculation of activity coefficients and potential phase separations in well-characterized organic-inorganic mixtures, AIOMFAC is certainly the better choice.

Specific comments

"p. 15306, line 13. The authors use parameters from Peng (2001), but the UNIFAC model of Raatikainen (2005), described for the same compound types (diacids and hydroxy acids), is based on a larger and more diverse data set. Is there a reason why the parameter set of Peng was preferred over that of Raatikainen?"

There is a simple reason for that: Raatikainen and Laaksonen (2005) modified their UNIFAC version by fitting interactions of $COOH$ and OH with water and CH_n to (new) literature data, but they also refitted the interactions between water and the CH_n main group, and between CH_n and $COOH$ to better represent their datasets. Hence, the refitted parameters are no longer compatible with those of standard UNIFAC, for which the interaction parameters between $CH_n \leftrightarrow H_2O$ and $CH_n \leftrightarrow COOH$ had been fitted using a much larger database including various types of functional groups (and not only CH_n, OH, and $COOH$). In AIOMFAC we retain the latter interaction parameters from the standard UNIFAC because they are very likely already well constrained and this approach also maintains better compatibility regarding all main groups used. In contrast to Raatikainen and Laaksonen (2005), Peng et al. (2001) only modified UNIFAC parameters concerning the interactions of $COOH \leftrightarrow H_2O$, $COOH \leftrightarrow OH$, and $H_2O \leftrightarrow OH$; however, the latter is not even used in AIOMFAC as we use UNIFAC-MP for $H_2O \leftrightarrow OH$ interaction parameters. Hence, the modified parameters of Peng et

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

al. (2001) can be integrated easily, improving the representation of dicarboxylic acids (and other multifunctional acids), by changing only a few interaction parameters. In addition, Table 5 of Raatikainen and Laaksonen (2005) shows that the Peng et al. (2001) UNIFAC is as good as their modified version for most of the data they compared the models to.

"p. 15307, line 3-5. The $CH_n^{(OH)}$ group leads to a better description of alcohols, polyols and sugars. But the authors should also mention other molecules with hydroxy groups. What about e.g. hydroxy acids and hydroxy ketones? For example, citric acid is within AIOMFAC described by parameters from original UNIFAC, UNIFAC-Peng and UNIFAC-MP. I could imagine that its description will be less optimal than by UNIFAC-Peng only."

As commented above, the difference of using the combination of standard UNIFAC, UNIFAC-Peng, and UNIFAC-MP in case of citric acid and other carboxylic acids with hydroxyl groups, instead of using standard UNIFAC + UNIFAC-Peng parameters only, was found to be small. The OH group interaction parameters in UNIFAC-Peng were fitted using a rather limited number of different compounds containing hydroxyl groups. Therefore, the use of UNIFAC-MP for the OH group interaction parameters is much better justified and the results of using the combination of all UNIFAC parameter sets as defined in AIOMFAC support this approach.

"p. 15313, Eq. (9) and following text. w_d^{init} is determined based on data type and temperature range, but how exactly? Did you apply some protocol or was it simply chosen by intuition?"

The rule was to distinguish between different data types and the covered temperature

range. LLE and SLE data close to 298 K were assigned an initial weighting of 1.0, $\gamma_{\pm}$ and a_w(bulk) data an initial weighting of 2.0, a_w(EDB) data 1.0, and VLE data an initial weighting of 0.5 (or less depending on temperature range). However, there are a number of exceptions: datasets showing large scatter or inconsistency with most other data were given lower weightings or were even set to zero. The dataset contributions to the objective function value after trial optimization iterations was analyzed and employed to identify possible inconsistencies among different datasets, possible mistakes in the dataset conversion or the implementation in the model (quality control), and to avoid that a few datasets completely dominate the parameter optimization (due to numerical issues or other unjustified reasons). Thus, the protocol was to use a defined weighting, but if it seemed necessary, to allow for modifications of those weightings based on experience and considerations regarding the overall model optimization.

"p. 15316, Eq. (11). C=C is also in the first series, while it has two carbon atoms. Shouldn't it be removed from this first series?
Same equation. Is it possible to locate a zero-point (referring to water) in both series?"

The C=C is shown in both series to establish a link between the two series. Although it contains two carbon atoms, its position in the first series is correct with respect to the Gibbs excess energy contribution. Water represents the zero point in both series regarding the value of $B_{k,i^\star}(I)$ for all I (for k = H$_2$O). This zero point is typically to the left, but in some cases slightly to the right of the COOH position in the polarity series. However, as this depends on both the ionic strength I as well as the ion type, no numerical fix point can be simply located with regards to the shown polarity series.

"p. 15319, Eq. (13). I am a bit surprised to see a difference of activity coefficients here, instead of a ratio (or a difference of their logarithms). Normally it is the logarithms of activity coefficients that are added, e.g. the functional group

activity coefficients in the UNIFAC model to obtain the molecular activity coefficient, or the weighted average of $\ln\gamma_i$ to obtain the Gibbs molar excess energy. Was a difference of logarithms also tried, and was the current approach found to be superior?"

Logarithms of different contributions to the molecular activity coefficients are added in models, because this is mathematically derived directly from the additivity of different Gibbs excess energy contributions, yet this has little to do with the use of a difference in Eq. (13). Whether the difference between the activity coefficients obtained from salt-containing and salt-free data is used for the quantities in the objective function or their ratio, will mainly influence the scaling ("weighting") of the contributions to the objective function with respect to deviations at different component mole fractions. There is no mathematical reason why a difference should not be used. The use of activity coefficient ratios had been considered, but we concluded that the difference works well for the parameter fitting.

"p. 15320, line 9. Can the "+1" term be derived mathematically, or was it artificially introduced?"

As described on page 15320, the term "+1" was introduced in the denominator of Eq. (8) (the obj. function) to normalize the salt-effect isolation treatment of VLE data. The introduction appears somewhat artificial, but the reason is mathematically motivated and is related to the point above concerning the use of a difference of activity coefficients in Eq. (13). Since at some compositions, the activity difference of Eq. (13) as well as corresponding sensitivity values can become very small, the denominator of Eq. (8) for that data point could become very small when calculated without the "+1" term, leading to a high objective function contribution of this point, even for relatively small differences between the calculated and measured quantities in the numerator. This would typically occur for components at compositions where

they are at high mole fractions, e.g., for water at $x_w > 0.9$, and where very little of the salt-effect is represented by the difference between calculated and experimentally derived quantities in the nominator. Rather, a combination of random measurement uncertainty and uncertainty related to the way the activity coefficients of corresponding salt-free solutions were derived using Eqs. (16) and (17) within the "salt-effect isolation treatment" could easily dominate the contributions to Eq. (8) at such compositions. Such undesired behavior is very effectively attenuated by the normalization term "+1" in the denominator, where this value is equal to an activity coefficient in an ideal solution. The additional "+1" term ensures that the significant contributions of a VLE dataset to the objective function value relate to the actual salt-effect and not to uncertainty artifacts.

"p. 15320, last line and first line of 15321. You could show the formula how to calculate y1 with AIOMFAC. Is for this data type $(x - y - T$ data without $p)$ no isolation of the salt-effect attempted?"

The data type VLE(x-y-T) (without total pressure given) in our database coincides with ternary mixtures containing monocarboxylic acids, such as propanoic acid, which associate in the gas phase to form significant amounts of dimers. Therefore, individual activity coefficients of water and the organic compound cannot be derived from the experimental data in the way it is achieved for complete VLE(x-y-p-T) data and no isolation of the salt-effect is attempted. However, the measured mole fraction of water in the gas phase, $y_1 = \frac{p_1}{p_1 + p_{2m} + 2p_{2d}}$ (Chueh et al., 1974), can be compared to the AIOMFAC calculation of y_1 when the gas-phase association equilibrium is considered. In the expression for y_1, p_1 is the partial pressure of water, while p_{2m} and p_{2d} are the partial pressures of carboxylic acid (component 2) monomer and dimer, respectively. The partial pressures of monomer and dimer are related by: $p_{2d} = (p_{2m})^2 K_2(T)$, where K_2 is the temperature dependent association equilibrium constant as given, e.g., by

Tsonopoulos and Prausnitz (1970). The expression to calculate y_1 with AIOMFAC is then:

$$y_1 = \frac{\gamma_1^{(x)} x_1 p_1^\circ}{\gamma_1^{(x)} x_1 p_1^\circ + \gamma_2^{(x)} x_2 p_{2m}^\circ + 2K_2 \left(\gamma_2^{(x)} x_2 p_{2m}^\circ \right)^2}. \tag{1}$$

Here, p_{2m}° is the saturation vapor pressure of the monomer of component 2 at T, calculated by solving the quadratic equation: $K_2(p_{2m}^\circ)^2 + p_{2m}^\circ - p_2^\circ = 0$, where p_2° is the total saturation vapor pressure of component 2 at T, obtained from the Antoine Equation with coefficients from the Landolt-Börnstein database (Dykyj et al., 2000). Hence the unknowns in Eq. (1) are the activity coefficients, which are calculated with AIOMFAC. The tolerance quantity of data type VLE(x-y-T) is set to $Q_{d,u}^{\text{tol}} = x^{\text{tol}}$.

"Section 4.3. It was not completely clear to me how the LLE data are taken into account:"

- *"p. 15323. line 21. "set to the activity value". I guess that from LLE data one only has the concentrations in both phases, not the activities. One only knows that the activities have to be the same in both phases. So is the "activity value" here obtained by using the experimental concentrations and the AIOMFAC calculated activity coefficient?"*

Yes, the activity values are calculated with AIOMFAC at the given experimental compositions (see p. 15323, line 8). The point we try to make on page 15323 is that the reference quantity, $Q_{d,u}^{\text{ref}}$, of a component can be either its calculated activity in phase α or in phase β. Whether phase α or phase β is the reference phase for the component depends on its activity sensitivities in the two phases, evaluated over all points of the dataset.

C7530

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- *"p. 15323, line 24. "rel. activity deviation". It is not clear to me how this quantity is defined. Is it defined as $\frac{a_j^\beta - a_j^\alpha}{a_j^\alpha + s_j^{a,\alpha} + s_j^{a,\beta}}$, where concentrations are experimental and activity coefficients are calculated by AIOMFAC?"*

  The "rel. activity deviation" is calculated using the model activities at experimental phase compositions in the data point expression of the objective function (Eq. 8). Hence, the expression given by the reviewer is correct when the reference phase of j is phase α (otherwise all α superscripts have to be switched to β and vice versa). (This will be made more clear in the revised version of the manuscript).

- *"p. 15323, line 23. The sensitivities for both phase activities are summed. Shouldn't it be $\sqrt{\left(s_j^{a,\alpha}\right)^2 + \left(s_j^{a,\beta}\right)^2}$ according to the error propagation rule for sums?"*

  The reviewer is right, $\sqrt{\left(s_j^{a,\alpha}\right)^2 + \left(s_j^{a,\beta}\right)^2}$ should have been used for $Q_{d,u}^{\text{tol}}$ in this context. Since in most cases the activity sensitivity of a component in one of the two phases is much larger than in the other, the error made by using the sum of the sensitivities is relatively small.

- *"p. 15323, line 25. "The absolute values of the relative activity deviations ... are calculated relative to". Shouldn't the first "relative" be scrapped here?"*

  Indeed, the sentence should read: "The absolute values of the activity deviations as plotted in Fig. 4 are calculated relative to the compound-specific reference

phase activities plus activity sensitivities, analogously to the calculation of their contributions to F_{obj}.".

"p. 15336, line 15. I think that also the molecule classes "hydroperoxides" and "nitrates" should be added as to be beyond the scope of UNIFAC, at least as long publicly accessible parameters are meant. Do the authors have any recommendations when such molecule classes are encountered? E.g. working by analogy (e.g. treat hydroperoxides as alcohols), or simply ignoring non-ideality for such compounds?"

We will add these classes. Organic nitrates are also mentioned along with organosulfates as potentially being formed in the condensed phase. When such molecule classes are required in calculations with UNIFAC or AIOMFAC, we recommend to use the UNIFAC groups and interaction parameters given by Compernolle et al. (2009). For the interaction parameters between ions and these main groups an analogy approach as suggested by Compernolle et al. (2009) could be used. Ignoring non-ideality for such compounds by setting unknown interactions coefficients in AIOMFAC to zero is not recommended and will lead to large deviations when inorganic ions and water are present.

"p. 15337, line 8 "However, it is at present not possible to provide a quantitative estimate of how well AIOMFAC will perform for mixtures that were not part of the database used for the parameter determination." Why is this not possible? One could for example split the dataset into two, using one part as the fitting set and the other part as the test set to test the predictive capacity of the model."

Du to the highly variable number of datasets available to constrain individual organic main group–ion interaction pairs, splitting the database into a reasonable fitting set

and a test set would be quite difficult. A more systematic approach would be to use all datasets except for one, to try the model predictability for that one dataset and then use this procedure one-by-one for all datasets. However, the point we want to make here is that this is at present not possible for technical reasons: fitting the 250 AIOMFAC parameters simultaneously with a global optimization algorithm and that many data points already causes high computational costs – looping over all datasets with the test set approach would take very long (more than a year). Another important point is that an approach using a test set could only reveal how the model predictability is for mixtures containing organic compounds that are eventually used in the overall parameter fit, but not how well AIOMFAC predictions will be for much larger and more complex molecular structures not present in any dataset of the database.

Minor comments

"p. 15329, line 13. "theses" should be "these" "

Fig. 6, caption. "(Chiavone-Filho and Rasmussen, 1993)" should be "Chiavone-Filho and Rasmussen (1993)"."

We will correct these points in the revised manuscript.

References

Chueh, C. F.: Predicting activity coefficients of multicomponent solutions containing formic acid, acetic acid, formaldehyde, acetaldehyde, and water, AIChE Symp. Ser., 70, 110–119, 1974.

Compernolle, S., Ceulemans, K., and Müller, J.-F.: Influence of non-ideality on condensation to aerosol, Atmos. Chem. Phys., 9, 1325–1337, doi:10.5194/acp-9-1325-2009, 2009.

Donahue, N. M., Epstein, S. A., Pandis, S. N., and Robinson, A. L.: A two-dimensional volatility basis set: 1. organic-aerosol mixing thermodynamics, Atmos. Chem. Phys., 11, 3303-3318, doi:10.5194/acp-11-3303-2011, 2011.

Fredenslund, A., Jones, R. L., and Prausnitz, J. M.: Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures, AIChE J., 21, 1086–1099, 1975.

Hansen, H. K., Rasmussen, P., Fredenslund, A., Schiller, M., and Gmehling, J.: Vapor–liquid-equilibria by UNIFAC group contribution. 5. Revision and extension, Ind. Eng. Chem. Res., 30, 2352–2355, 1991.

Marcolli, C. and Peter, Th.: Water activity in polyol/water systems: new UNIFAC parameterization, Atmos. Chem. Phys., 5, 1545–1555, doi:10.5194/acp-5-1545-2005, 2005.

Peng, C., Chan, M. N., and Chan, C. K.: The hygroscopic properties of di-carboxylic and multi functional acids: Measurements and UNIFAC predictions, Environ. Sci. Technol., 35, 4495–4501, doi:10.1021/es0107531, 2001.

Raatikainen, T. and Laaksonen, A.: Application of several activity coefficient models to waterorganic- electrolyte aerosols of atmospheric interest, Atmos. Chem. Phys., 5, 2475–2495, doi:10.5194/acp-5-2475-2005, 2005.

Tsonopoulos, C. and Prausnitz, J. M.: Fugacity Coefficients in Vapor-Phase Mixtures of Water and Carboxylic Acids, Chem. Eng. J., 1, 273–278, doi:10.1016/0300-9467(70)85014-6, 1970.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper