

Prediction of Infinite-Dilution Activity Coefficients Using UNIFAC and COSMO-SAC Variants

Renan P. Gerber and Rafael de P. Soares*

Departamento de Engenharia Química, Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Rua Engenheiro Luis Englert, s/n, Bairro Farroupilha, CEP 90040-040, Porto Alegre, RS, Brazil

Infinite-dilution activity coefficients (IDAC) can be used to predict, for example, the behavior of liquid–liquid equilibrium or to determine parameters for excess Gibbs free energy expressions like NRTL, Wilson, or UNIQUAC. For systems where limited (or no) experimental data are available, predictive tools as the well-known UNIFAC variations can be very helpful. An alternative approach is to use methods based on conductor-like screening model (COSMO). In this work, we have tested a modified UNIFAC model with a recent parameter set and three variations of the COSMO-SAC model. To compare the models, a database with a total of 748 IDAC experimental points was assembled with data from the literature and is available as Supporting Information. For nonaqueous solutions, the IDAC logarithm absolute average deviation for UNIFAC was 0.28, while it was 0.48 for a COSMO-SAC with parameters adjusted in this work. On the other hand, for aqueous systems, a COSMO-SAC with optimized parameters led to a deviation of 0.81 against 1.47 for the UNIFAC model tested.

Introduction

The activity coefficient measures the degree of nonideality of a compound in solution. The greatest degree of nonideality is at infinite dilution, when a single solute molecule is completely surrounded by solvent. The infinite-dilution activity coefficient (IDAC - γ^∞) can be used to predict, for example, the behavior of liquid–liquid equilibrium or to determine parameters for excess Gibbs free energy expressions like NRTL, Wilson, or UNIQUAC.

For systems where limited (or no) experimental data are available, predictive tools such as the well-known UNIFAC variations can be very helpful.¹ These methods take advantage of the fact that the free energy of molecules in solution is (to a considerable degree) additive.² Despite the undoubted merits of these models, they need large sets of experimental data for the derivation of the subgroup and group interaction parameters and suffer from the inability to distinguish between isomers.

An alternative approach is to use methods based on the conductor-like screening model (COSMO). These models, which are based on computational quantum mechanics, allow us to predict thermo-physical properties without any experimental data. A few input parameters are necessary and need to be estimated only once for all substances.^{3,4} COSMO for real solvents (COSMO-RS), introduced in 1995 by Klamt,² was the first model in this category. The required COSMO calculation has been implemented in many quantum chemistry software packages such as Gaussian,^{5,6} TURBOMOLE,⁷ MOPAC,⁸ DMol3,⁹ and GAMESS.¹⁰ In 2002, Lin and Sandler¹¹ proposed a COSMO-RS variation: COSMO segment activity coefficient (COSMO-SAC). Later, Gmehling and Gmehling¹² developed another modification called COSMO-RS(OI). At the current stage of development, the COSMO-based models are quite close to reaching the point of becoming a practical chemical engineering utility.¹³ It can be expected that these models will become a valuable supplement for the traditional group contribution methods in physical chemistry and chemical engineering.

There have been a number of critical evaluations of the ability of methods at predicting infinite dilution activity coefficients, γ^∞ . In Thomas and Eckert,¹⁴ Gmehling et al.,¹⁵ Kontogeorgis et al.,¹⁶ Kontogeorgis et al.,¹⁷ and Voutsas et al.,¹⁸ several models have been tested for their accuracy in the prediction of γ^∞ in nonaqueous mixtures. For this kind of mixture, the values of γ^∞ are generally low, with an average value below 10. Voutsas and Tassios¹ assessed the ability to predict γ^∞ of nonelectrolyte mixtures, with emphasis on aqueous ones where γ^∞ can be larger than 10¹⁰, using several different models (mostly variants of UNIFAC). Later, Putnam et al.¹³ tested the COSMO-RS performance in the prediction of infinite dilution activity coefficients and compared the results to those obtained by Voutsas and Tassios.¹ The authors have found promising results for aqueous systems.

In this work, we have tested a modified UNIFAC model and COSMO-SAC variants with respect to their accuracy in the prediction of infinite-dilution activity coefficients. The modified UNIFAC implementation tested here uses the recent parameter matrix developed by Horstmann et al.¹⁹ to be used with the PSRK mixing rule. Three variations of COSMO-SAC were also tested. The first one has σ -profiles (detailed later in this Article) from DMol3 and parameter settings as described by Lin and Sandler,¹¹ referred to hereinafter as COSMO-SAC-D^{lit}. The second one, also based on DMol3, but with parameters optimized in this work, is referred to just as COSMO-SAC-D. The third one was coupled with σ -profiles generated in this work using MOPAC,⁸ referred to hereinafter as COSMO-SAC-M. All models use the same COSMO-SAC implementation present in the computer program JCOSMO, but differ in the σ -profile and/or parameter settings. JCOSMO is a Java implementation of the COSMO-SAC model developed by the authors. It is open-source, with code available via the Internet at <http://code.google.com/p/jcosmo>.

The models were compared using an expanded γ^∞ database, with data from Voutsas and Tassios¹ and He and Zhong.²⁰ We have also identified erroneous values in the database of Voutsas and Tassios. The expanded and corrected database is available as Supporting Information along with predicted IDAC values for 12 models: three versions of UNIFAC,^{21–23} four modified

* To whom correspondence should be addressed. Tel.: +55 51 33083528. Fax: +55 51 33083277. E-mail: rafael@enq.ufrgs.br.

UNIFAC implementations,^{24–27} and COSMO-RS.¹³ Besides these predicted values taken from Putnam et al.,¹³ we have also included the values predicted with a modified UNIFAC with a recent parameter matrix;¹⁹ COSMO-SAC-D^{lit}; COSMO-SAC-M; and COSMO-SAC-D. The assembled database was also used to optimize some of the universal model parameters for COSMO-SAC-M and COSMO-SAC-D.

Models

Modified UNIFAC. The group contribution method UNIFAC is a Gibbs excess free energy (G^E) model, which allows the prediction of liquid-phase activity coefficients γ_i in nonelectrolyte systems as a function of temperature and composition. The activity coefficient is calculated as the sum of a combinatorial and a residual part:

$$\ln \gamma_i = \ln \gamma_i^{\text{comb}} + \ln \gamma_i^{\text{res}} \quad (1)$$

The combinatorial part ($\ln \gamma_i^{\text{comb}}$) accounts for differences in the size and shape of the molecules (contribution of the excess entropy), and the residual part ($\ln \gamma_i^{\text{res}}$) accounts mainly for the effects that arise from energetic interactions between groups (contribution of the excess enthalpy).²⁸

For the UNIFAC predictions, an implementation from the thermodynamic package VRTherm²⁹ was used. This package contains a modified UNIFAC implementation with the parameter matrix published by Horstmann et al.¹⁹ with three parameters for the temperature-dependent term Ψ_{mn} :

$$\Psi_{mn} = \exp\left(-\frac{a_{mn} + b_{mn}T + c_{mn}T^2}{T}\right) \quad (2)$$

where a_{mn} , b_{mn} , and c_{mn} are the group interaction parameters.

COSMO-SAC. COSMO-RS, introduced by Klamt,² was the first extension of a dielectric continuum-solvation model to liquid-phase thermodynamics. COSMO-SAC is a variation of COSMO-RS, proposed by Lin and Sandler.¹¹

In the COSMO-SAC model, the activity coefficient can be defined as the result of two contributions:¹¹

$$\ln \gamma_{i/s} = \frac{\beta(\Delta G_{i/s}^{\text{res}} - \Delta G_{i/i}^{\text{res}})}{RT} + \ln \gamma_{i/s}^{\text{SG}} \quad (3)$$

The first is the difference between the free energies of restoring the charges around the solute molecule in solution S and restoring the charges in a pure liquid i , scaled by an empirical factor β . The second contribution is the Staverman–Guggenheim (SG) combinatorial term, which improves the calculations for the cavity-formation free energy:

$$\ln \gamma_{i/s}^{\text{SG}} = \ln\left(\frac{\phi_i}{x_i}\right) + \frac{z}{2}q_i \ln\left(\frac{\theta_i}{\phi_i}\right) + l_i - \frac{\phi_i}{x_i} \sum_j x_j l_j \quad (4)$$

where $\phi_i = r_i x_i / \sum_j r_j x_j$ is the normalized volume fraction; $\theta_i = q_i x_i / \sum_j q_j x_j$ is the normalized surface-area fraction; $l_i = (z/2)(r_i - q_i) - (r_i - 1)$; z is the coordination number; x_i is the mole fraction; $r_i = V_i/r$ and $q_i = A_i/q$ are the normalized volume and surface-area, respectively; A_i is the cavity surface area and V_i is the cavity volume; and q and r are universal parameters of the model.

The empirical scaling factor β was introduced in this work and is another universal parameter of the model. As discussed later, the introduction of this parameter significantly improved the experimental data fit.

The restoring free energy of the solute can be obtained by:

$$\frac{\Delta G_{i/s}^{\text{res}}}{RT} = n_i \sum_{\sigma_m} p_i(\sigma_m) \ln \Gamma_s(\sigma_m) \quad (5)$$

where $\Gamma_s(\sigma_m)$ is the activity coefficient for a segment of charge density σ and is given by:

$$\ln \Gamma_s(\sigma_m) = -\ln\left\{\sum_{\sigma_n} p_s(\sigma_n) \Gamma_s(\sigma_n) \exp\left[\frac{-\Delta W(\sigma_m, \sigma_n)}{RT}\right]\right\} \quad (6)$$

$$\Delta W(\sigma_m, \sigma_n) = \left(\frac{\alpha'}{2}\right)(\sigma_m + \sigma_n)^2 + c_{hb} \max[0, \sigma_{acc} - \sigma_{hb}] \times \min[0, \sigma_{don} + \sigma_{hb}] \quad (7)$$

$$\alpha' = f_{\text{pol}} \left(\frac{0.3a_{\text{eff}}^{3/2}}{\epsilon_0} \right) \quad (8)$$

where α' is the constant for the misfit energy; $\epsilon_0 = 2.395 \times 10^{-4}$ (e² mol)/(kcal Å) is the permittivity of a vacuum; f_{pol} is the polarization factor; c_{hb} is a constant for hydrogen bonding; σ_{hb} is the σ -value cutoff for hydrogen bonding; and σ_{acc} and σ_{don} are the larger and smaller values of σ_m and σ_n .

The probability of finding an element with a charge density σ (the σ -profile) in a pure substance i is represented by:

$$p_i(\sigma) = \frac{n_i(\sigma)}{n_i} = \frac{A_i(\sigma)}{A_i} \quad (9)$$

where $n_i(\sigma)$ is the number of segments with charge density σ , A_i is the total cavity surface area, and $A_i(\sigma)$ is the total surface area of all of the segments with a particular charge density σ .

The total number of surface segments in a molecule is $n_i = A_i/a_{\text{eff}}$, where $a_{\text{eff}} = \pi r_{\text{eff}}^2$ is the standard segment surface area and r_{eff} is the averaging radius, one of the adjustable parameters in this method.

One of the basic assumptions of the COSMO-RS method, also inherited by COSMO-SAC, is that all averaged surface segments are independent. Thus, the σ -profile of an ensemble of molecules $p(\sigma)$ is simply composed of the σ -profiles of its components:

$$p(\sigma) = \sum_{i=1}^N x_i p_i(\sigma) \quad (10)$$

where x_i is the mole fraction of the substance i .

The calculation to obtain the σ -profile of a molecule starts with the computation of the ideal screening charge density distribution over the molecular surface using COSMO. The result is a surface divided into segments, each one with its own area and charge density. Once COSMO-SAC requires all segments with identical area, the surface charge density distribution from the COSMO output (σ^*) needs to be averaged to find an effective surface charge density (σ) using, for instance, the following equation:

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{\text{eff}}^2}{r_n^2 + r_{\text{eff}}^2} \exp\left(\frac{d_{mn}^2}{r_n^2 + r_{\text{eff}}^2}\right)}{\sum_n \frac{r_n^2 r_{\text{eff}}^2}{r_n^2 + r_{\text{eff}}^2} \exp\left(\frac{d_{mn}^2}{r_n^2 + r_{\text{eff}}^2}\right)} \quad (11)$$

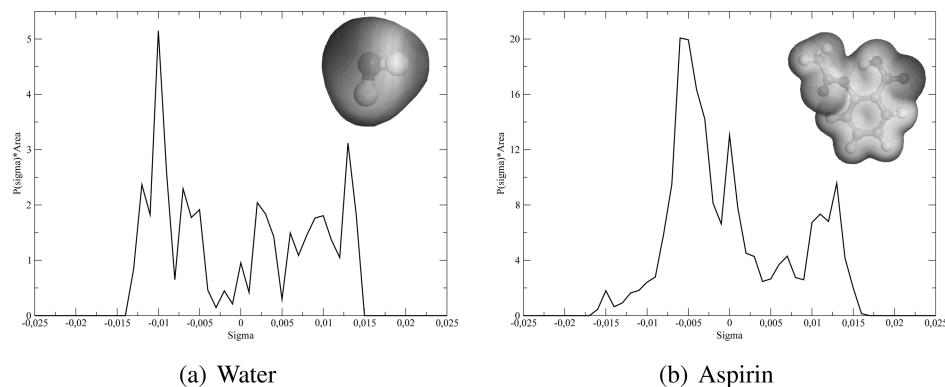


Figure 1. Sigma profiles for water (a) and aspirin (b) generated in this work using MOPAC, three-dimensional images from Mullins et al.³¹ using DMol3.

Table 1. Literature Parameter Values for COSMO-SAC-D^{lit} and Optimized Parameters for COSMO-SAC-M and COSMO-SAC-D

	β	c_{hb} [kcal \AA^4]/ mol \cdot e ²]	σ_{hb} [e/ \AA^2]	q [\AA^2]
COSMO-SAC-D ^{lit}	1.00	85 508	0.00840	79.53
COSMO-SAC-M	1.30	21 364	0.00499	33.62
COSMO-SAC-D	1.12	25 580	0.00595	80.83

where d_{mn} is the distance between segment m and n ; r_n is the radius of the actual surface segment; and r_{eff} is the averaging radius, an adjustable parameter set to 0.81764 \AA by Wang et al.³⁰

Finally, the three-dimensional geometric charge density distribution is projected onto a two-dimensional histogram, which is known as the σ -profile, $p_i(\sigma)$. This is exemplified for two different substances in Figure 1.

JCOSMO. In 2008, the computational program known as JCOSMO was implemented by the authors. The program JCOSMO consists of an implementation of the COSMO-SAC model in the Java programming language. JCOSMO is not a quantum chemistry package; it does not execute COSMO calculations, but can compute the activity coefficient of mixtures using σ -profiles determined previously. Currently, JCOSMO comes with the σ -profiles provided by Mullins et al.³¹ Further, JCOSMO is open-source, with code available via the Internet at <http://code.google.com/p/jcosmo>. In summary, the current features of the program are database with 1432 compounds (σ -profile using DMol3 by Mullins et al.³¹); can predict activity coefficients for liquid mixtures with any number of compounds using the COSMO-SAC model; can predict the infinite dilution activity coefficients for binary mixtures; a graphical interface for plotting the activity coefficient profiles; and a graphical interface for plotting the σ -profiles.

MOPAC σ -Profiles. Although Mullins et al.³¹ have made available a large σ -profile database, with data for 1432 substances, it is limited to the selected compounds. In this work, we have investigated the performance of COSMO-SAC when coupled with σ -profiles from molecular orbital package (MOPAC), referred to as the COSMO-SAC-M model. MOPAC is a semiempirical quantum chemistry program based on Dewar and Thiel's NDDO approximation.⁸

It is important to note that the COSMO-SAC model requires consistency between the quantum calculations and the model parameter settings.³² Some of the adjustable parameters for the COSMO-SAC method can be seen in Table 1. In this work, the models referred as COSMO-SAC-D^{lit} and COSMO-SAC-D are based on DMol3 σ -profiles from Mullins et al.³¹ For

COSMO-SAC-D^{lit}, the parameters were kept as published by the authors, while optimized parameters are used in COSMO-SAC-D.

Mu et al.³³ compared σ -profiles produced by Gaussian 03, TURBOMOLE, and DMol3. MOPAC differs from these packages in the level of sophistication. MOPAC is based on semiempirical methods and tends to be less precise than the other packages, which are based entirely on theory from first principles. On the other hand, the approximations considered in semiempirical methods enable it to reduce the computing time by orders of magnitude. For instance, MOPAC can perform a COSMO computation with geometry optimization in 0.12 s for WATER. If GAMESS is used (which compares to DMol3 in complexity) for a similar computation using the configuration flags as described by Wang et al.³² plus geometry optimization, the computation time is 61.7 s. For TOLUENE, GAMESS takes 621 min to finish a COSMO computation with geometry optimization, while MOPAC takes only 1.11 s. These tests were performed using a single processor of an Intel(R) Core(TM) 2 Quad CPU with 3.2 Gb of memory running Ubuntu Linux 10.04. Thus, the use of MOPAC in this comparative study also aims at shedding some light on the merits of the approximate solution produced by semiempirical packages when coupled with COSMO-based activity coefficient models.

When properly configured, MOPAC is able to generate the required COSMO information in one of its output files (with the .COS extension). Using this file, we can produce the σ -profile using the averaging procedure described earlier. In the Appendix, a sample MOPAC input file is given.

In Figure 2, σ -profiles produced in this work using MOPAC are compared to those made available by Mullins et al.³¹ using DMol3. As was already observed by Mu et al.,³³ different quantum mechanical packages produce different σ -profiles. In general, the profiles produced using MOPAC are slightly narrower than those produced by DMol3. In this case, a new reparametrization is mandatory.

Results and Discussion

Experimental Database. The IDAC database from Voutsas and Tassios¹ contains 500 experimental data points, 386 for nonaqueous systems and 114 for aqueous systems. In this work, we have assembled an expanded database by adding 250 new experimental data points for aqueous systems from He and Zhong.²⁰ Further, by inspecting the outliers in predicted versus experimental charts, we have identified two erroneous points in the database from Voutsas and Tassios:¹ cyclopentane/water at 298 K and methyl *n*-propyl ketone/water at 298 K. These

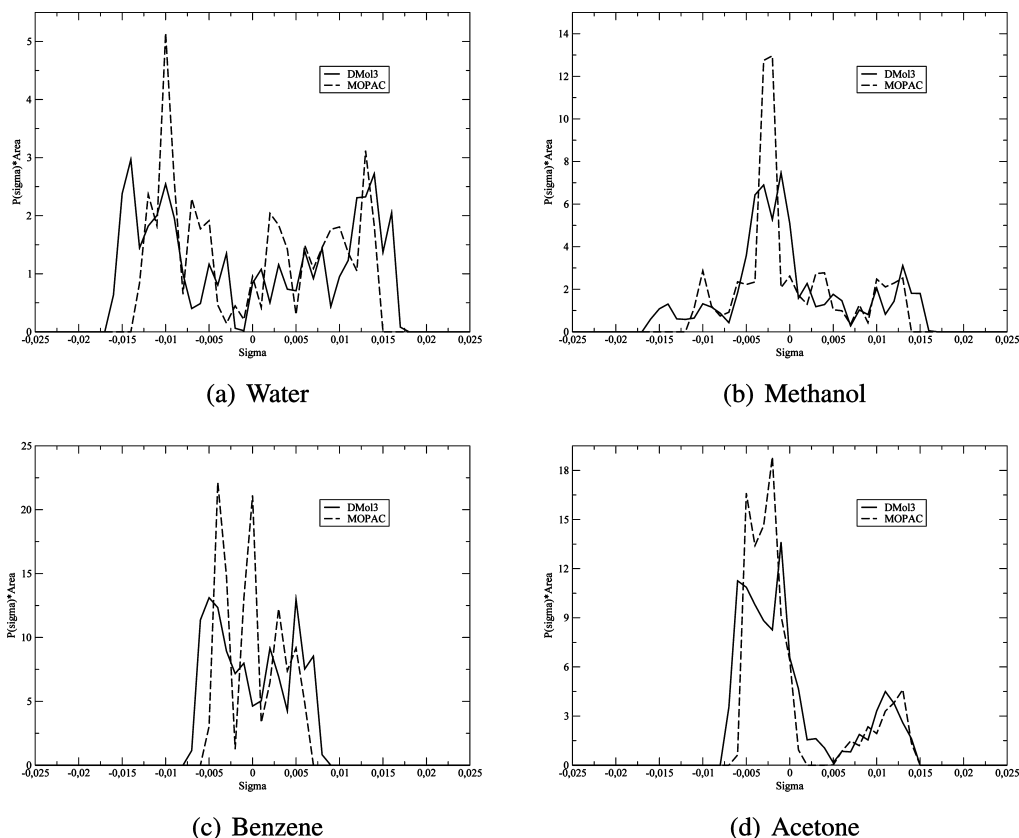


Figure 2. Sigma profiles generated with MOPAC and DMol3³¹ for some substances.

points were replaced with similar data found in the work of He and Zhong.²⁰

The expanded and corrected database with a total of 748 data points is available as Supporting Information.

Parameter Optimization. As expected, we experienced a poor performance when the MOPAC σ -profiles were used with the parameters adjusted for another package (DMol3). From Figure 2, it is also clear that the σ -profiles generated by MOPAC are narrower than those produced by DMol3. This will directly impact in the hydrogen-bond sigma cutoff (σ_{hb}). The profiles also differ in several aspects, which are difficult to correlate with the model parameters. Thus, a sensitivity analysis was executed to discover which are the more relevant parameters for the IDAC prediction.

The criteria used in the sensitive analysis as well as in the parameter optimization were the IDAC logarithm absolute average deviation (ln AAD):

$$\ln \text{AAD} = \frac{1}{\text{NP}} \sum_{i=1}^{\text{NP}} |\ln \gamma_i^{\text{exp}} - \ln \gamma_i^{\text{calc}}| \quad (12)$$

where NP is the total number of points.

To select which model parameters should be optimized, we inspected the changes in ln AAD when applying small variations in the parameters. The σ -profile averaging radius as well as all COSMO parameters (atom specific radius, etc.) were not considered here because these cannot be adjusted when using the σ -profiles already computed by Mullins et al.³¹ We have verified that the model is insensitive to the volume normalization factor r ; this is actually a well-known result (e.g., Chapter 5 of Michelsen and Møllerup³⁴) but not clearly stated in the COSMO-SAC literature. The area normalization factor q and the coordination number z are strongly correlated, and we

decided to optimize only q . Finally, we ended up with a set of four universal parameters to be adjusted. These parameters were optimized using a direct search method.³⁵ Experimental data for the following families were not considered in the parameter tuning process: amine, carboxylic acid, ether, and ketone. This is because COSMO-based models usually have a poor performance with this kind of substance.^{30,32,33} For COSMO-SAC-M, the ln AAD was 2.73 when using literature parameters and 0.85 after the optimization. Likewise, for COSMO-SAC-D, the ln AAD was 1.19 when using literature parameters and 0.54 after the optimization. The resulting parameter values along with the literature parameters are shown in Table 1.

IDAC Predictions. To quantitatively compare IDAC models, the deviation error among the estimated and the experimental data can be calculated either with ln AAD, eq 12, or with:

$$\% \text{AAD} = \frac{100}{\text{NP}} \sum_{i=1}^{\text{NP}} \frac{|\gamma_{i,\text{exp}} - \gamma_{i,\text{calc}}|}{\gamma_{i,\text{exp}}} \quad (13)$$

Several works in the literature use the percent absolute average deviation (% AAD), eq 13, to assess model performance.^{1,13,33} The problem with % AAD is that very small contributions are accounted if the model underestimates the value (especially when the experimental value is large). This is not the case if the ln AAD, eq 12, is used. Another reason to prefer ln AAD is its direct relation to the excess Gibbs free energy. Thus, ln AAD is used instead of % AAD in numerical comparisons in this work.

Using the extended database described in the previous section, we have compared the performance of a modified UNIFAC, COSMO-SAC-D^{lit}, COSMO-SAC-M, and COSMO-SAC-D. For the COSMO-SAC models, the JCOSMO program was used. For UNIFAC predictions, the implementation from the thermodynamic package VRTherm²⁹ was used.

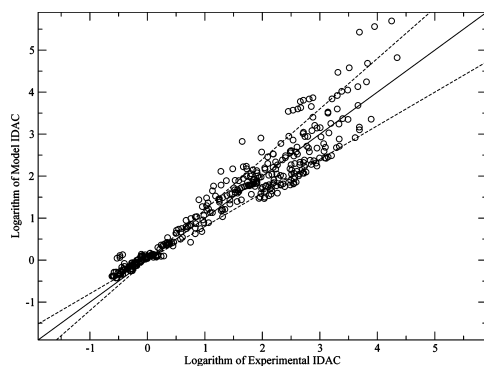
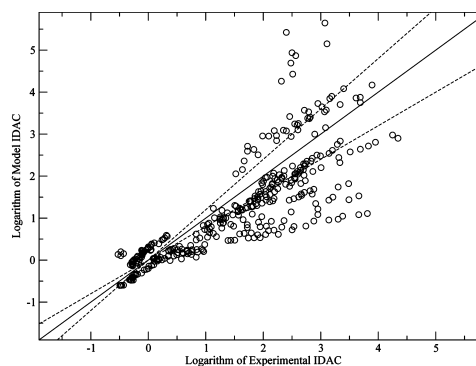
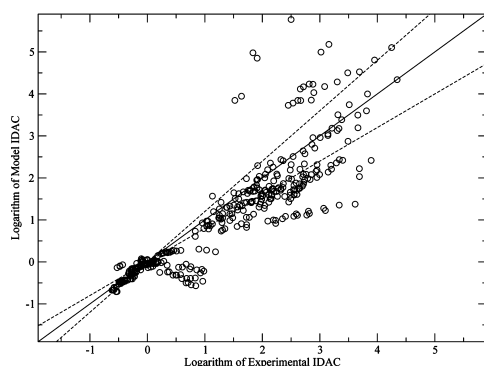
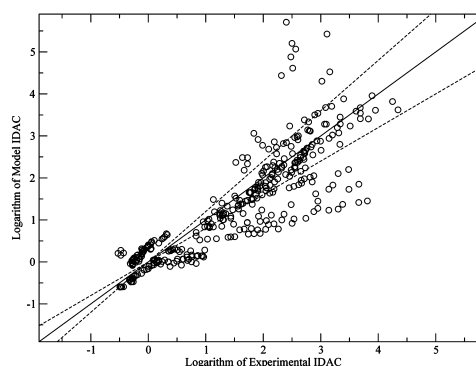
(a) UNIFAC, $NP = 374$ (b) COSMO-SAC-D^{lit}, $NP = 361$ (c) COSMO-SAC-M, $NP = 384$ (d) COSMO-SAC-D, $NP = 361$

Figure 3. Experimental IDAC logarithm versus model IDAC logarithm for nonaqueous systems. The number of points (NP) varies according to model data availability; dashed lines represent 20% departure.

Table 2. IDAC Logarithm Absolute Average Deviation (ln AAD) for Solutes in Nonaqueous Systems

solute	solvent	UNIFAC ²⁶		COSMO-RS ¹³		UNIFAC		COSMO-SAC-D ^{lit}		COSMO-SAC-M		COSMO-SAC-D	
		NP	ln AAD	NP	ln AAD	NP	ln AAD	NP	ln AAD	NP	ln AAD	NP	ln AAD
alcohol	alkane	16	0.14	14	0.89	16	0.33	16	0.42	16	1.18	16	0.27
alcohol	cycloalkane	15	0.21	15	0.80	15	0.20	15	0.47	15	1.28	15	0.19
alkane	alcohol	20	0.15	16	0.19	20	0.56	20	0.94	20	0.40	20	0.40
alkane	alkyl halide	5	0.03	5	0.12	5	0.12	5	0.25	5	0.25	5	0.35
alkane	carboxylic acid	4	0.28	4	0.84	4	0.17	4	0.50	4	2.52	4	0.87
alkane	amine	7	0.16	6	1.43	7	0.26	7	1.82	7	0.34	7	0.76
alkane	ketone	21	0.12	21	0.28	21	0.17	21	0.51	21	0.29	21	0.12
alkane	phenol	24	0.12	24	0.36	24	0.58	24	0.67	24	0.88	24	0.47
alkene	amine	18	0.21	15	1.42	18	0.28	18	1.83	18	0.12	18	0.77
alkyl halide	alkane	42	0.03	22	0.11	42	0.07	39	0.16	42	0.10	39	0.28
amine	alkane	11	0.10	11	0.66	11	0.93	11	1.34	11	1.24	11	0.97
aromatic	alkane	76	0.04	28	0.47	68	0.18	54	0.22	74	0.08	54	0.11
carboxylic acid	alkane	4	0.15	0		4	0.22	4	2.11	4	1.99	4	2.21
carboxylic acid	cycloalkane	4	0.43	0		4	0.45	4	1.91	4	4.82	4	1.95
cycloalkane	alcohol	6	0.08	6	0.16	6	0.12	6	0.42	6	0.32	6	0.22
cycloalkane	alkyl halide	5	0.03	2	0.01	5	0.04	5	0.06	5	0.08	5	0.10
cycloalkane	amine	1	0.88	1	1.76	1	0.71	1	1.92	1	0.93	1	1.29
cycloalkane	carboxylic acid	4	0.30	4	0.31	4	0.36	4	0.25	4	1.71	4	0.41
cycloalkane	phenol	28	0.22	27	0.26	28	0.48	28	0.42	28	0.57	28	0.27
cycloalkene	amine	12	0.11	6	0.96	6	0.20	12	1.12	12	0.12	12	0.37
ketone	alcohol	34	0.07	34	0.42	34	0.26	34	0.54	34	0.86	34	0.51
ketone	alkane	29	0.14	28	0.15	29	0.14	29	0.45	29	0.19	29	0.33
weighted average		386	0.12	290	0.46	374	0.28	361	0.59	384	0.56	361	0.48

Nonaqueous Systems. Our database contains 384 points for nonaqueous solutions. The results for the IDAC predictions for nonaqueous systems are in Figure 3. It is clear that the UNIFAC model tested performed better for this category of mixtures, even when compared to COSMO-SAC models with optimized parameters.

In Table 2, the quality of the model predictions is numerically compared. The number of points (NP) reported differ between models because we have missing model information for some

substances. As a result, we have tested 374 systems for the UNIFAC model, 361 systems for the COSMO-SAC-D model, and 384 systems for the COSMO-SAC-M model.

Inspecting the values in Table 2, we will again conclude that, in general, the modified UNIFAC implementation tested performs better than the COSMO-SAC models for nonaqueous solutions. The worst results for COSMO-SAC models were obtained for carboxylic acid systems. The COSMO-SAC-D performed better than COSMO-SAC-M with very few excep-

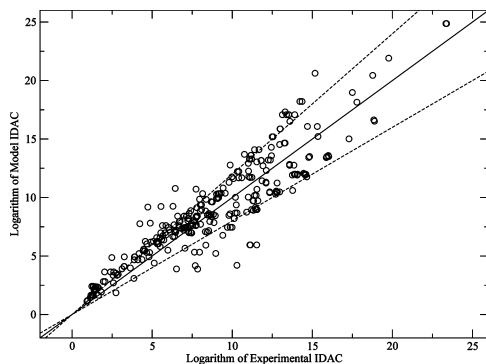
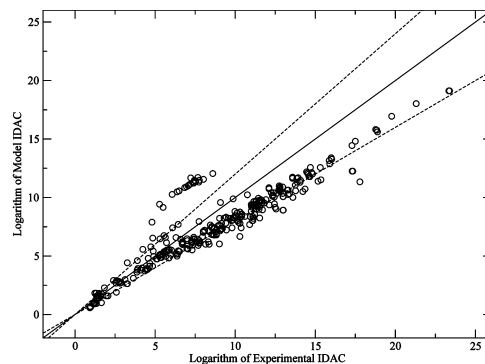
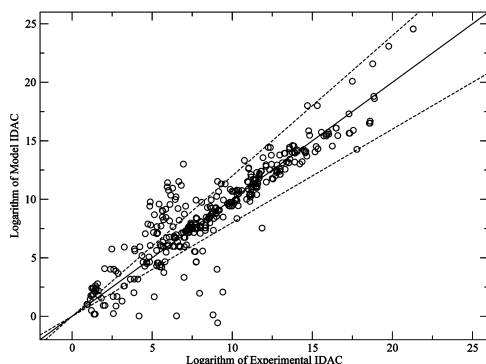
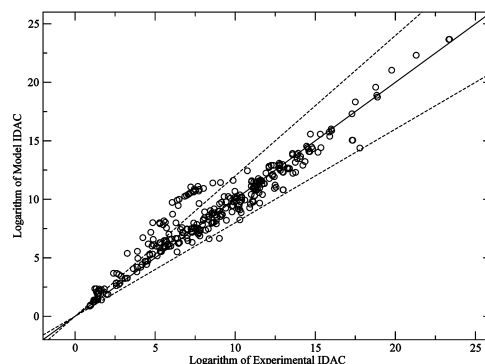
(a) UNIFAC, $NP = 294$ (b) COSMO-SAC-D^{lit}, $NP = 311$ (c) COSMO-SAC-M, $NP = 359$ (d) COSMO-SAC-D, $NP = 311$

Figure 4. Experimental IDAC logarithm versus model IDAC logarithm for aqueous systems. The number of points (NP) varies accordingly to model data availability; dashed lines represent 20% departure.

Table 3. IDAC Logarithm Absolute Average Deviation ($\ln AAD$) for Solutes in Aqueous Systems and Water in Solvents

solute type	UNIFAC ²⁶		COSMO-RS ¹³		UNIFAC		COSMO-SAC-D ^{lit}		COSMO-SAC-M		COSMO-SAC-D	
	NP	$\ln AAD$	NP	$\ln AAD$	NP	$\ln AAD$	NP	$\ln AAD$	NP	$\ln AAD$	NP	$\ln AAD$
alcohol	16	1.03	7	0.53	39	1.17	40	1.74	58	2.69	40	0.96
aldehyde	7	0.55	3	0.30	15	0.69	15	0.69	15	0.82	15	0.19
alkane	15	4.32	6	0.72	39	2.29	39	2.32	39	0.38	39	0.22
alkene	0		0		14	1.10	13	1.99	15	0.31	13	0.68
alkyne	0		0		4	1.23	6	2.34	6	0.56	6	1.23
alkyl halide	9	0.94	7	0.34	32	1.34	33	1.52	39	2.15	33	0.46
aromatic	6	0.45	6	0.10	23	1.65	24	2.14	24	0.40	24	0.50
aryl halide	0		0		4	4.52	5	3.02	5	0.90	5	1.52
multiring aromatics	8	2.15	6	2.01	18	3.19	18	3.46	33	0.84	18	1.28
carboxylic acid	9	0.93	6	0.64	11	0.33	11	1.78	11	0.92	11	0.67
cycloalkane	4	3.69	4	1.06	11	1.34	11	1.69	11	0.47	11	0.24
cycloalkene	0		0		3	0.91	5	2.03	7	0.32	5	0.70
ether	0		0		9	3.10	12	0.81	14	1.99	12	2.09
ester	0		0		22	1.25	20	0.40	23	1.69	20	0.40
ketone	5	0.55	5	0.89	17	1.09	24	0.78	24	0.55	24	0.42
water ^a	35	1.41	35	0.31	35	0.46	35	2.42	35	0.41	35	2.01
weighted average	112	1.68	83	0.57	294	1.47	311	1.75	359	1.18	311	0.81

^a Water at infinite dilution in other solvents.

tions. Thus, the approximations present in MOPAC (when compared to more sophisticated packages) indeed impact in the model performance. Although on average UNIFAC performed better, it is interesting to note that COSMO-SAC-D performed better than UNIFAC in a significant number of families.

Aqueous Systems. The model deviations for aqueous systems tend to be larger than those for nonaqueous systems. This is because in aqueous systems it is easy to find an IDAC larger than 10^{10} . Figure 4 and Table 3 show that for this kind of system, on average, COSMO-SAC-D performs better than the UNIFAC model tested in this work. The largest deviation for

COSMO-SAC-D was observed with ether systems followed by water as a solute.

From Figure 4, we can also conclude that the COSMO-SAC models work especially well for large degrees of nonideality ($\ln \gamma^\infty > 10$). Comparing the results of COSMO-SAC-M and COSMO-SAC-D, we again conclude that the approximations present in MOPAC impact on the IDAC prediction using COSMO-SAC models.

Comparison with Literature Results. Tables 2 and 3 also contain literature results for COSMO-RS¹³ and another modified UNIFAC.²⁶ Regarding the COSMO-RS¹³ implementation, large

Table 4. IDAC Logarithm Absolute Average Deviation (ln AAD): Literature Data for a Modified UNIFAC²⁶ and COSMO-RS;¹³ Computed in This Work for a Modified UNIFAC, COSMO-SAC-D^{lit}, COSMO-SAC-M, and COSMO-SAC-D

model	nonaqueous		aqueous	
	NP	ln AAD	NP	ln AAD
UNIFAC ²⁶	386	0.12	114	1.80
COSMO-RS ¹³	290	0.46	85	0.57
UNIFAC	374	0.28	294	1.47
COSMO-SAC-D ^{lit}	361	0.59	311	2.05
COSMO-SAC-M	384	0.56	359	1.29
COSMO-SAC-D	361	0.48	311	0.81

deviations are present when working with carboxylic acids and amines. The same is true for all COSMO-based implementations tested, but COSMO-SAC-D performed better with amines. This is similar for alcohols in alkanes and cycloalkanes, where COSMO-RS¹³ has showed a deviation larger than 0.8, while for COSMO-SAC-D a deviation smaller than 0.3 was observed.

COSMO-based models should also be used with care when dealing with mixtures containing ketones and alcohols. Although relative small deviations are observed for this kind of mixture, the models always underestimate the activity coefficient.

Regarding aqueous systems, the worst performance for COSMO-RS¹³ was with multiring aromatics, followed by cycloalkane and ketone. For these three families, a better performance was observed when using COSMO-SAC-D. Large deviations were observed in systems with ether in water and water as a solute for COSMO-SAC-D.

In a direct comparison, COSMO-SAC-D was superior to COSMO-RS¹³ with the following families: alcohol/alkane; alcohol/cycloalkane; alkane/amine; alkane/ketone; alkene/amine; aromatic/alkane; cycloalkane/amine; cycloalkene/amine; aldehyde/water; alkane/water; multiring aromatics/water; cycloalkane/water; and ketone/water.

In Table 4, we have a more general comparison. In this table, literature results for COSMO-RS¹³ and another modified UNIFAC²⁶ are also included. From this table, we can see that, for nonaqueous solutions, COSMO-RS performed slightly better than the COSMO-SAC-D with a deviation of 0.46 (NP = 290) against 0.48 (NP = 361). Yet the smaller number of points considered in the COSMO-RS evaluation makes this result not decisive. The results also show that, on average, the modified UNIFAC implementations currently perform better than COSMO-based methods for the IDAC prediction of nonaqueous solutions.

Regarding aqueous systems, we can see that COSMO-RS performed much better than all other models, but the results are available only for a small number of points (NP = 85). For aqueous solutions, COSMO-SAC-D performed better than the modified UNIFAC implementations. We can also see that for this kind of system, the modified UNIFAC with the parameter matrix from Horstmann et al.¹⁹ superseded the previous version from Gmehling et al.²⁶

Conclusions

In this work, we have tested the accuracy in the prediction of infinite-dilution activity coefficients (IDAC – γ^∞) of a modified UNIFAC model and three variations of COSMO-SAC model: COSMO-SAC-D^{lit} with σ -profiles from DMol3 and literature parameters; COSMO-SAC-M with σ -profiles generated by MOPAC and parameters optimized in this work; and COSMO-SAC-D with σ -profiles from DMol3 and parameters optimized in this work. The UNIFAC implementation tested here uses a recent parameter matrix from Horstmann et al.¹⁹ All COSMO-SAC models use the same implementation present

in the computer program JCOSMO, but differ in the σ -profile and/or parameter settings. JCOSMO is a Java implementation of the COSMO-SAC model developed by the authors. It is open-source, with code available at <http://code.google.com/p/jcosmo>. As experimental data we used the IDAC database from Voutsas and Tassios¹ expanded with data from He and Zhong.²⁰ This expanded database was used to compare the models and to optimize some of the universal model parameters for COSMO-SAC-M and COSMO-SAC-D.

Regarding the prediction of infinite-dilution activity coefficients for nonaqueous solutions, in general UNIFAC performed better than the COSMO-based models. For the systems considered, the UNIFAC implementation tested has shown IDAC logarithm absolute average deviations (ln AAD) ranging from 0.04 (cycloalkane–alkyl halide systems) to 0.93 (amine–alkane systems). For a total of 374 experimental data points (NP) of nonaqueous systems, the UNIFAC ln AAD was 0.28. The literature shows even better results for another modified UNIFAC implementation,²⁶ with a deviation of only 0.12 for 386 points. The deviations for COSMO-SAC-M (NP = 384) and COSMO-SAC-D (NP = 361) were 0.56 and 0.48, respectively. This is similar to literature¹³ results of a COSMO-RS implementation with a deviation of 0.46 for NP = 290.

For aqueous systems, COSMO-SAC-D performed better than the other models tested in this study. The ln AAD for this model, for a total of 311 experimental points, was 0.81, while for UNIFAC (NP = 294) and COSMO-SAC-M (NP = 359) deviations were 1.47 and 1.29, respectively. Results from the literature¹³ show an even better performance for COSMO-RS with a deviation of only 0.57, but this was achieved with a limited number of experiments (NP = 85).

With very few exceptions, the COSMO-SAC-D predictions were significantly better than those obtained with COSMO-SAC-M. Thus, we can conclude that the simplifications considered in the semiempirical package MOPAC, in contrast to more sophisticated packages, impact on the IDAC prediction using COSMO-SAC models. On the other hand, a reparametrization of the semiempirical methods is also possible and will be investigated in future work.

Acknowledgment

This work was partially supported by PROPESQ-UFRGS-Brasil.

Supporting Information Available: The computer program JCOSMO is available, with sources included, via the Internet at <http://code.google.com/p/jcosmo>.

An expanded database with 748 experimental data points (386 points for nonaqueous and 362 for aqueous systems) along with the predicted values using 12 different models is available. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Appendix: MOPAC Input File

A MOPAC input file configured for COSMO calculations (water as example) follows:

```

EPS = 999.0 COSWRT RSOLV = 1.3 AM1 VDW(:H = 1.3:C =
2.0:N = 1.83:O = 1.72:F = 1.72:S = 2.16:P = 2.12:Cl = 2.05:Br =
2.16:I = 2.32) GNORM = 0.1 RELSCF = 0.1 WATER, ID: C7732185
O 0.000000 1 0.000000 1 0.000000 1 0 0 0
H 0.980209 1 0.000000 1 0.000000 1 1 0 0
H 0.980138 1 104.534696 1 0.000000 1 1 2 0

```

Nomenclature

List of Symbols

α' = constant for the misfit energy, kcal $\text{\AA}^4/\text{mol e}^2$
 ΔG^{res} = free energy of surface charge restoration, kcal/mol
 $\Delta W(\sigma_m, \sigma_n)$ = exchange energy, kcal/mol
 $\Gamma(\sigma_m)$ = segment activity coefficient
 γ^∞ = activity coefficient at infinite dilution
 γ_{is}^{SG} = Staverman–Guggenheim combinatorial contribution to the activity coefficient
 γ_i = activity coefficient of the substance i
 γ_i^{comb} = UNIFAC combinatorial term
 γ_i^{res} = UNIFAC residual term
 ϕ_i = normalized volume fraction
 Ψ_{mn} = UNIFAC group-interaction parameter between groups n and m
 σ = surface-segment charge-density distribution, $\text{e}/\text{\AA}^2$
 σ_n^* = surface-charge density for segment n from the COSMO output, $\text{e}/\text{\AA}^2$
 σ_{hb} = sigma cutoff for hydrogen bonding, $\text{e}/\text{\AA}^2$
 θ_i = normalized surface-area fraction
 A_i = total cavity surface area, \AA^2
 $A_i(\sigma)$ = surface area of all segments with a surface charge density σ , \AA^2
 a_{eff} = effective surface segment surface area, \AA^2
 a_{mn} = UNIFAC group-interaction parameter between groups n and m , K
 b_{mn} = UNIFAC group-interaction parameter between groups n and m
 c_{hb} = hydrogen-bonding constant, kcal $\text{\AA}^4/\text{mol e}^2$
 c_{mn} = UNIFAC group-interaction parameter between groups n and m , K^{-1}
 d_{mn} = distance between surface segment m and n , \AA
 f_{pol} = polarization factor
 $n_i(\sigma)$ = number of segments with a surface charge density of σ
 $p(\sigma)$ = sigma profile of a mixture
 $p_i(\sigma)$ = sigma profile of the pure substance i
 q_i = normalized surface-area parameter
 R = ideal gas constant
 r_i = normalized volume parameter
 r_n = effective radius of surface segment n , assuming circular surface segments, \AA
 r_{eff} = surface-segment averaging radius (adjustable parameter)
 T = temperature, K
 V_i = total cavity volume, \AA^3
 x_i = mole fraction of component i in the liquid phase
 z = coordination number

Literature Cited

- (1) Voutsas, E. C.; Tassios, D. P. Prediction of infinite-dilution activity coefficients in binary mixtures with UNIFAC. A critical evaluation. *Ind. Eng. Chem. Res.* **1996**, *35*, 1438–1445.
- (2) Klamt, A. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, 2224–2235.
- (3) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. Refinement and parametrization of COSMO-RS. *J. Phys. Chem. A* **1998**, *102*, 5074–5085.
- (4) Klamt, A.; Eckert, F. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. *Fluid Phase Equilib.* **2000**, *172*, 43–72.
- (5) Truong, T. N.; Stefanovich, E. V. A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity. *Chem. Phys. Lett.* **1995**, *240*, 253–260.
- (6) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (7) Schafer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F. COSMO implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187–2193.
- (8) MOPAC2009, James J. P. Stewart, Stewart Computational Chemistry, Version 10.124L web: <http://OpenMOPAC.net>.
- (9) Andzelm, J.; Kolmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995**, *103*, 9312–9320.
- (10) Baldrige, K.; Klamt, A. First principles implementation of solvent effects without outlying charge error. *J. Chem. Phys.* **1997**, *106*, 6622–6633.
- (11) Lin, S. T.; Sandler, S. I. A priori phase equilibrium prediction from a segment contribution solvation model. *Ind. Eng. Chem. Res.* **2002**, *41*, 899–913.
- (12) Grensemann, H.; Gmehling, J. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Ind. Eng. Chem. Res.* **2005**, *44*, 1610–1624.
- (13) Putnam, R.; Taylor, R.; Klamt, A.; Eckert, F.; Schiller, M. Prediction of infinite dilution activity coefficients using COSMO-RS. *Ind. Eng. Chem. Res.* **2003**, *42*, 3635–3641.
- (14) Thomas, E. R.; Eckert, C. A. Prediction of limiting activity coefficients by a modified separation of cohesive energy density model and UNIFAC. *Ind. Eng. Chem. Process Des. Dev.* **2002**, *23*, 194–209.
- (15) Gmehling, J.; Li, J.; Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **2002**, *32*, 178–193.
- (16) Kontogeorgis, G. K.; Coutsikos, P.; Tassios, D.; Fredenslund, A. Improved models for the prediction of activity coefficients in nearly athermal mixtures: Part I. Empirical modifications of free-volume models. *Fluid Phase Equilib.* **1994**, *92*, 35–66.
- (17) Kontogeorgis, G. M.; Nikolopoulos, G. I.; Fredenslund, A.; Tassios, D. P. Improved models for the prediction of activity coefficients in nearly athermal mixtures Part II. A theoretically-based GE-model based on the van der Waals partition function. *Fluid Phase Equilib.* **1997**, *127*, 103–121.
- (18) Voutsas, E.; Kalospiros, N. S.; Tassios, D. A combinatorial activity coefficient model for symmetric and asymmetric mixtures. *Fluid Phase Equilib.* **1995**, *109*, 1–15.
- (19) Horstmann, S.; Jabloniec, A.; Krafczyk, J.; Fischer, K.; Gmehling, J. PSRK group contribution equation of state: comprehensive revision and extension IV, including critical constants and α -function parameters for 1000 components. *Fluid Phase Equilib.* **2005**, *227*, 157–164.
- (20) He, J.; Zhong, C. A QSPR study of infinite dilution activity coefficients of organic compounds in aqueous solutions. *Fluid Phase Equilib.* **2003**, *205*, 303–316.
- (21) Hansen, H. K.; Rasmussen, P.; Fredenslund, A.; Schiller, M.; Gmehling, J. Vapor-liquid equilibria by UNIFAC group contribution. 5. Revision and extension. *Ind. Eng. Chem. Res.* **1991**, *30*, 2352–2355.
- (22) Hansen, H. K.; Coto, B.; Kuhlmann, B. UNIFAC with Linearly Temperature-Dependent Group-Interaction Parameters, Technical Report, Institut for Kemiteknik, DTH: Lyngby, 1992.
- (23) Magnussen, T.; Rasmussen, P.; Fredenslund, A. UNIFAC parameter table for prediction of liquid-liquid equilibria. *Ind. Eng. Chem. Process Des. Dev.* **1981**, *20*, 331–339.
- (24) Larsen, B. L.; Rasmussen, P.; Fredenslund, A. A modified UNIFAC group-contribution model for prediction of phase equilibria and heats of mixing. *Ind. Eng. Chem. Res.* **1987**, *26*, 2274–2286.
- (25) Bastos, J. C.; Soares, M. E.; Medina, A. G. Infinite dilution activity coefficients predicted by UNIFAC group contribution. *Ind. Eng. Chem. Res.* **1988**, *27*, 1269–1277.
- (26) Gmehling, J.; Li, J.; Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Ind. Eng. Chem. Res.* **1993**, *32*, 178–193.
- (27) Hooper, H. H.; Michel, S.; Prausnitz, J. M. Correlation of liquid-liquid equilibria for some water-organic liquid systems in the region 20–250 °C. *Ind. Eng. Chem. Res.* **1988**, *27*, 2182–2187.
- (28) Jakob, A.; Grensemann, H.; Lohmann, J.; Gmehling, J. Further development of modified UNIFAC (Dortmund): Revision and extension 5. *Ind. Eng. Chem. Res.* **2006**, *45*, 7924–7933.
- (29) VRTEch, 2005; www.vrtech.com.br.
- (30) Wang, S.; Sandler, S. I.; Chen, C.-C. Refinement of COSMO-SAC and the applications. *Ind. Eng. Chem. Res.* **2007**, *46*, 7275–7288.
- (31) Mullins, E.; Oldland, R.; Liu, Y. A.; Wang, S.; Sandler, S. I.; Chen, C.-C.; Zwolak, M.; Seavey, K. C. Sigma-profile database for using COSMO-based thermodynamic methods. *Ind. Eng. Chem. Res.* **2006**, *45*, 4389–4415.

(32) Wang, S.; Lin, S.-T.; Watanasiri, S.; Chen, C.-C. Use of GAMESS/COSMO program in support of COSMO-SAC model applications in phase equilibrium prediction calculations. *Fluid Phase Equilib.* **2009**, 276, 37–45.

(33) Mu, T.; Rarey, J.; Gmehling, J. Performance of COSMO-RS with sigma profiles from different model chemistries. *Ind. Eng. Chem. Res.* **2007**, 46, 6612–6629.

(34) Michelsen, M. L.; Mollerup, J. M. *Thermodynamic Models: Fundamentals and Computational Aspects*, 2nd ed.; Tie-Line Publications: Holte, Denmark, 2007.

(35) Wright, M. H. Direct Search Methods: Once Scorned, Now Respectable. In *Proceedings of the 1995 Biennial Conference on Numerical Analysis*; Griffiths, D. F., Ed.; Addison Wesley Longman Ltd.: Reading, MA, 1996.

Received for review December 14, 2009

Revised manuscript received June 22, 2010

Accepted June 28, 2010

IE901947M