

# ASSESSING THE RELIABILITY OF PREDICTIVE ACTIVITY COEFFICIENT MODELS FOR MOLECULES CONSISTING OF SEVERAL FUNCTIONAL GROUPS

R. P. Gerber and R. P. Soares\*

Departamento de Engenharia Química, Escola de Engenharia, Universidade Federal do Rio Grande do Sul,  
Phone: + (55) (51) 3308-3528, Fax: + (55) (51) 3308-3277, Rua Engenheiro Luis Englert, s/n,  
Bairro Farroupilha, CEP: 90040-040, Porto Alegre - RS, Brazil.  
E-mail: rafael@enq.ufrgs.br  
E-mail: rpgerber@enq.ufrgs.br

(Submitted: September 26, 2011 ; Revised: November 16, 2011 ; Accepted: November 21, 2011)

**Abstract** - Currently, the most successful predictive models for activity coefficients are those based on functional groups such as UNIFAC. In contrast, these models require a large amount of experimental data for the determination of their parameter matrix. A more recent alternative is the models based on COSMO, for which only a small set of universal parameters must be calibrated. In this work, a recalibrated COSMO-SAC model was compared with the UNIFAC (Do) model employing experimental infinite dilution activity coefficient data for 2236 non-hydrogen-bonding binary mixtures at different temperatures. As expected, UNIFAC (Do) presented better overall performance, with a mean absolute error of 0.12 ln-units against 0.22 for our COSMO-SAC implementation. However, in cases involving molecules with several functional groups or when functional groups appear in an unusual way, the deviation for UNIFAC was 0.44 as opposed to 0.20 for COSMO-SAC. These results show that COSMO-SAC provides more reliable predictions for multi-functional or more complex molecules, reaffirming its future prospects.

**Keywords:** COSMO-SAC; IDAC; MOPAC; Quantum mechanical; UNIFAC.

## INTRODUCTION

Reliable prediction of thermodynamic properties of substances in solution is the key for the design and optimization of many industrial processes. For separation processes in particular, the prediction of activity coefficients is crucial.

Currently, the most successful predictive models for activity coefficients are group contributions methods such as UNIFAC and modified UNIFAC (Grensemann and Gmehling, 2005). However, these models require a large amount of experimental data in order to fit the required group interaction parameter matrix. A recent alternative is the models

based on COSMO (COnductor-like Screening MOdel), which use results from quantum mechanical calculations (QM) instead of experimental data and only a small set of universal parameters must be calibrated. The first model in this category was COSMO-RS (Klamt, 1995).

In order to expand its use and to improve the model performance, the UNIFAC parameter matrix is being constantly updated. This is accomplished by reviewing the current parameters and by adding new groups and subgroups. Frequently, especially for complex substances, very specific and less comprehensive groups are needed for proper prediction. As an example we can mention the

\*To whom correspondence should be addressed

This article is an extended version of a work presented at CBTermo-2011 - VI Brazilian Congress of Applied Thermodynamics 2011, Salvador, Bahia, Brasil.

*lactone* group. It can be represented by the existing groups, in this case an ester, but due to poor performance a new group was introduced by the UNIFAC Consortium. Unfortunately, for each new group added a large amount of new parameters should to be estimated (interaction with all the existing groups). This makes it very costly to introduce new groups and, hence, the model correction for these complex substances is difficult. For these situations, COSMO-based models may provide more accurate results than the UNIFAC method, as will be shown later.

In this work the COSMO-RS variant known as COSMO-SAC (Segment Activity Coefficient) (Lin and Sandler, 2002) was coupled with QM calculations generated by the semi-empirical package MOPAC (Stewart, 2009). The open-source package JCosmo, developed by our group (Gerber and Soares, 2010), was used.

Among the parameters that need to be calibrated in the QM package and the COSMO-SAC model itself are the atomic radii, the solvent radius, parameters for electrostatic interactions and hydrogen bonding. In this study, in order to reduce the number of parameters to be calibrated and, given to the complexity of hydrogen-bonding systems, only mixtures known to not form hydrogen bonds were considered. In the present work, the model parameters were recalibrated using an extended experimental database of activity coefficients at infinite dilution (IDAC).

The predictions of the recalibrated model were compared with an implementation of the UNIFAC (Do) model with literature parameters (Jakob *et al.*, 2006), referred to hereafter as just UNIFAC. Comparisons were carried out using the mean absolute error (defined below) of the prediction of the natural logarithm of the IDAC. An experimental set comprising 2236 binary mixtures, involving 203 different substances at different temperatures, was used for parameter calibration and model comparison. This dataset is also available as supplementary material at <http://www.enq.ufrgs.br/labs/lvpp>.

## MODELS

The group contribution method UNIFAC and the COSMO-based models are Gibbs excess free energy ( $G^E$ ) models, which allow the prediction of liquid-phase activity coefficients  $\gamma_i$  as a function of temperature and composition (Fredenslund *et al.*, 1975). In both families of models the activity

coefficient is calculated as the sum of a combinatorial and a residual part:

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

The combinatorial part accounts for differences in the size and shape of the molecules and the residual part accounts mainly for the effects that arise from differences in the free energy interactions.

## UNIFAC

In this work, for the UNIFAC predictions, an implementation in the thermodynamic package VRTherm (VRTech, 2005) was used. This package contains a modified UNIFAC (Gmehling *et al.*, 1993) implementation with the parameter matrix published by Jakob *et al.* (2006). Three parameters are used for the temperature-dependent term  $\Psi_{mn}$ :

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

where  $a_{mn}$ ,  $b_{mn}$  and  $c_{mn}$  are the group interaction parameters for the group pair m-n, recalling that the parameter matrix is not symmetrical (*e.g.*  $a_{mn} \neq a_{nm}$ ).

## COSMO-SAC

COSMO-RS, introduced by Klamt (1995), was the first extension of a dielectric continuum-solvation model to liquid-phase thermodynamics. COSMO-SAC is a variation of COSMO-RS, proposed later by Lin and Sandler (2002).

The COSMO-SAC combinatorial contribution used in this work is similar to that present in the modified UNIFAC model, as suggested by Soares (2011):

$$\ln \gamma_i^{\text{comb}} = \ln \varphi' + 1 + \varphi' + \frac{z}{2} q_i \left( \ln \left( \frac{\varphi_i}{\theta_i} \right) + 1 - \frac{\varphi_i}{\theta_i} \right) \quad (3)$$

where  $\varphi_i = r_i x_i / \sum_j r_j x_j$  is the volume fraction;  $\theta_i = A_i x_i / \sum_j A_j x_j$  is the normalized surface-area fraction;  $z$  is the coordination number, usually taken as 10;  $x_i$  is the mole fraction;  $r_i = V_i / r$   $A_i$  is the cavity surface area and  $V_i$  is the cavity volume;  $r$  is

a universal parameter of the model;  $\varphi' = r_i^p x_i / \sum_j r_j^p x_j$  and  $p$  is another parameter of the model.

The energetic contribution is based on a theory of contacting surfaces. First each molecule is assumed to be inside of a cavity immersed in a perfect conductor. In this hypothetical situation, apparent surface charges can be computed using the COSMO method. This is a numerical method; therefore, the surface needs to be discretized into *surface segments* or *patches* before the actual numerical solution. Typical results for the apparent surface charges are depicted in Figure 1(a). In essence, the model takes into account the energetic difference between two molecule surface segments in contact with respect to their initial contact with a perfect conductor. The Figure 1(b) represents one possible contact. Finally, we can imagine that each molecule, initially immersed in a conductor, is incrementally put in contact with other molecules until the real solution state is achieved (where there is no conductor in between).

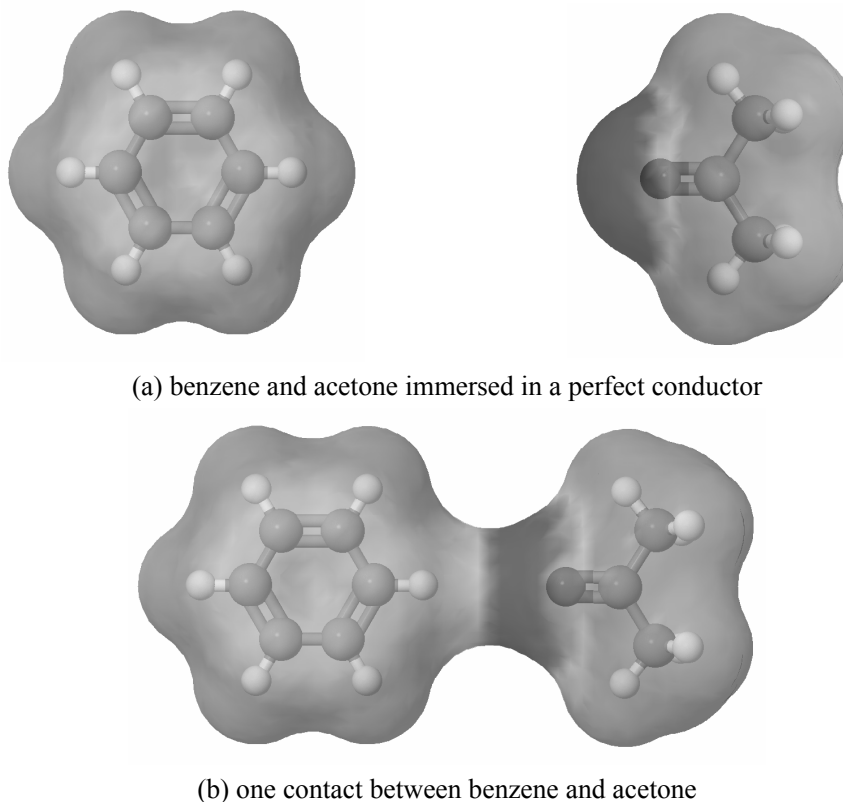
Of course, in the real solution there are infinite

possible arrangements and a statistical thermodynamic treatment is necessary. The first step for this treatment is the projection of the three-dimensional apparent surface charges onto a two-dimensional histogram, which is known as the  $\sigma$ -profile,  $p_i(\sigma)$ .

Because COSMO-SAC requires all segments to have an identical area, the surface charge density distribution from the COSMO output ( $\sigma^*$ ) needs to be averaged to find an effective surface charge density ( $\sigma$ ) using, for instance, the following equation:

$$\sigma_m = \frac{\sum_n \sigma_n^* \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)}{\sum_n \frac{r_n^2 r_{av}^2}{r_n^2 + r_{av}^2} \exp\left(\frac{d_{mn}^2}{r_n^2 + r_{av}^2}\right)} \quad (4)$$

where  $d_{mn}$  is the distance between segments  $m$  and  $n$ ;  $r_n$  is the radius of the actual surface segment;  $r_{av}$  is the averaging radius.



**Figure 1:** Representation of two molecules in a conductor (a) and in contact (b).

The probability of finding an element with a charge density  $\sigma$  in a molecule  $i$  is represented by:

$$p_i(\sigma) = \frac{n_i(\sigma) A_i(\sigma)}{n_i A_i} \quad (5)$$

where  $n_i(\sigma)$  is the number of segments with charge density  $\sigma$ ,  $A_i$  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  $\sigma$ . The total number of surface segments in a molecule is  $n_i = A_i/a_{\text{eff}}$ , where  $a_{\text{eff}} = \pi r_{\text{av}}^2$  is the standard segment surface area and  $r_{\text{av}}$  is the averaging radius, one of the adjustable parameters studied in this paper. It is usual in the literature to find two values for the averaging radius (Grensemann & Gmehling, 2005; Klamt *et al.*, 1998), one used to determine the surface area ( $r_{\text{eff}}$ ) and another for the  $\sigma$ -profile averaging procedure ( $r_{\text{av}}$ ). In this work,  $r_{\text{av}}$  will be used in both calculations.

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  $\sigma$ -profile of an ensemble of molecules  $p(\sigma)$  is simply composed of the  $\sigma$ -profiles of its components:

$$p(\sigma) = \sum_{i=1}^N x_i p_i(\sigma) \quad (6)$$

Thus the activity coefficient can be calculated by:

$$\ln \gamma_{i/s} = \frac{\beta (\Delta G_{i/s}^{*\text{res}} - \Delta G_{i/i}^{*\text{res}})}{RT} + \ln \gamma_i^{\text{comb}} \quad (7)$$

where the residual part, the first term, 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  $\beta$  (Gerber and Soares, 2010). We note here that without the factor  $\beta$ , a reasonable experimental data fit is only possible by choosing different values for  $r_{\text{av}}$  and  $r_{\text{eff}}$ . This term can be viewed as the contribution of excess enthalpy.

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

$$\frac{\Delta G_{i/s}^{*\text{res}}}{RT} = n_i \sum_{\sigma_m} p_i(\sigma_m) \ln \Gamma_s(\sigma_m) \quad (8)$$

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

$$\ln \Gamma_s(\sigma_m) = -\ln \left\{ \sum_{\sigma_n} p_s(\sigma_n) \Gamma_s(\sigma_n) \exp \left[ \frac{-\Delta W(\sigma_m, \sigma_n)}{RT} \right] \right\} \quad (9)$$

$$\Delta W(\sigma_m, \sigma_n) = \left( \frac{\alpha'}{2} \right) (\sigma_m + \sigma_n)^2 + c_{\text{hb}} \max[0, \sigma_{\text{acc}} - \sigma_{\text{hb}}] \min[0, \sigma_{\text{don}} + \sigma_{\text{hb}}] \quad (10)$$

where  $\alpha'$  is the constant for the misfit energy;  $c_{\text{hb}}$  is a constant for hydrogen bonding;  $\sigma_{\text{hb}}$  is the  $\sigma$ -value cutoff for hydrogen bonding; and  $\sigma_{\text{acc}}$  and  $\sigma_{\text{don}}$  are the larger and smaller values of  $\sigma_m$  and  $\sigma_n$ .

The second term on the right hand side of Eq. (10) is a simple approach to take into account hydrogen bond formation. The hydrogen bonding contribution has been widely studied in the literature and various formulations have been proposed (*e.g.* Wang *et al.*, 2007; Hsieh *et al.*, 2010). Unfortunately, no COSMO-based formulation currently available can properly describe the complexity of hydrogen bonding in a general way. Thus, in this study, in order to reduce the number of parameters to be calibrated and to avoid this complexity, only mixtures known to not form hydrogen bonds were considered. The  $c_{\text{hb}}$  parameter could then be set to zero as well as  $\sigma_{\text{hb}}$ .

The COSMO-SAC predictions were performed by modifying the computational program known as JCosmo, previously implemented by our group and described elsewhere (Gerber and Soares, 2010).

## PARAMETERIZATION

In the following sections, several aspects of the COSMO-SAC parametrization are investigated. When comparing the model performances with respect to experimental data, the following equation was considered:

$$\ln \text{IDAC}_{\text{error}} = \frac{1}{N} \sum_{i=1}^N \left| \ln \gamma_i^{\infty \text{exp}} - \ln \gamma_i^{\infty \text{model}} \right| \quad (11)$$

### Quantum Chemical Method

As described previously, the first step for any COSMO-based model is the determination of the apparent surface charges. This step can be accomplished by a variety of tools, ranging from semi-empirical methods, as implemented in MOPAC

(Stewart, 2009), to more sophisticated first-principle ones, implemented in many quantum chemistry software packages such as Gaussian (Barone and Cossi, 1998; Truong and Stefanovich, 1995), TURBOMOLE (Schäfer *et al.*, 2000), DMol3 (Andzelm *et al.*, 1995) and GAMESS (Baldrige and Klamt, 1997). The approximations considered in semi-empirical methods enable them to reduce the computing time by orders of magnitude (Gerber and Soares, 2010). Further, for large molecules, a direct solution may only be feasible with semi-empirical methods. It is therefore interesting to have the COSMO-SAC model fine tuned for a semi-empirical package. The package selected for this task was MOPAC, which is free for academic use.

Among the alternative methods available in MOPAC we have preselected AM1 (Dewar *et al.*, 1985) and RM1 (Rocha *et al.*, 2006). The semi-empirical quantum chemical method AM1 was introduced in the 80's and, although semi-empirical molecular orbital models still are of limited accuracy, it presents consistently good results and time-tested reliability. A more recent method, proposed in 2005, is a reparameterization of AM1 named RM1 (Recife Model 1). The properties used in the parameterization procedure were: heats of formation, dipole moments, ionization potentials and geometric variables (bond lengths and angles).

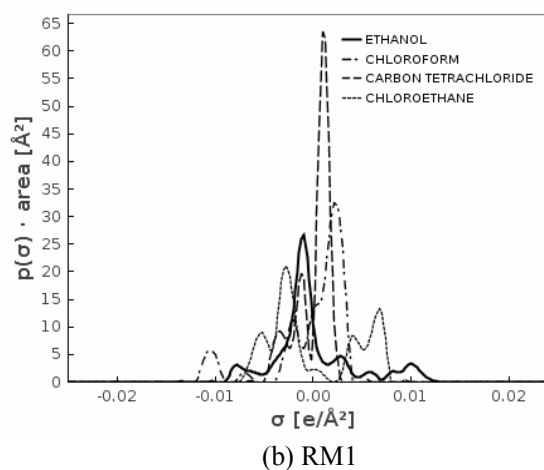
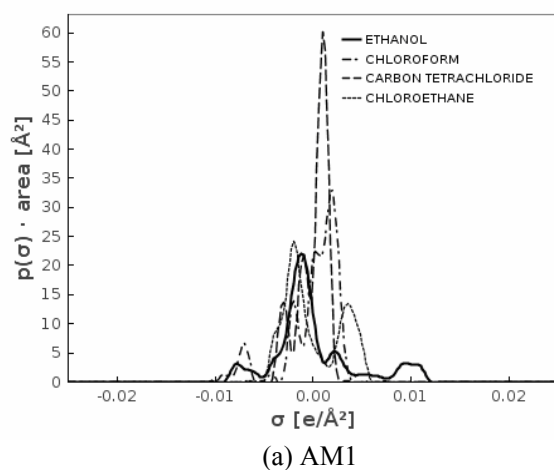
The apparent surface charges determined by COSMO calculations are strongly related to the dipole moments. Thus, this was the main property analyzed to select the method to be used in the present work. The average error in the prediction of the dipole moment by the AM1 and RM1 models for three groups of atoms are presented in Table 1

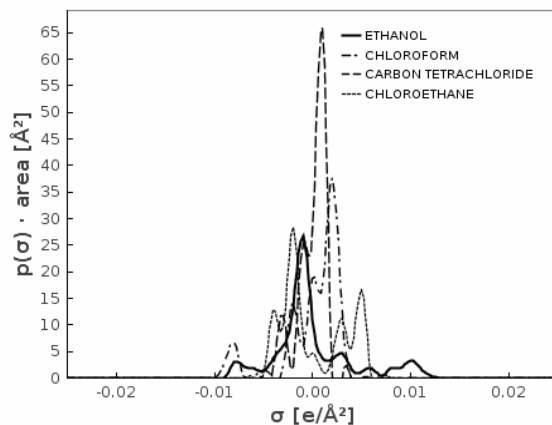
(Rocha *et al.*, 2006). RM1 is the best for most of the atoms but for F, Cl, Br, and I AM1 has a smaller average error. In order to take into account the best fit for the dipole moment, a combination of the two sets of parameters was used, called here POA1. POA1 uses the parameters of RM1 for C, H, N, O, P, and S atoms and the parameters of AM1 for F, Cl, Br, and I atoms.

**Table 1: Average errors ( $\Sigma_i 400(\mu_i^{\text{cal}} - \mu_i^{\text{exp}})^2$ ) for the dipole moment (D) for AM1, RM1 and POA1 semiempirical models for three groups of atoms, from Rocha *et al.* (2006).**

Atoms	NP	AM1	RM1	POA1
C, H, N, and O	59	0.26	0.23	0.23
P and S	16	0.74	0.49	0.49
F, Cl, Br, and I	52	0.37	0.42	0.37

Although the combination of parameters from different models might be dangerous, better results were indeed observed with POA1 when compared with the original models AM1 and RM1. Further, a qualitative analysis can be performed by examining the sigma profile of several substances. In Figure 2, the sigma profiles of three substances containing the Cl atom (chloroform, carbon tetrachloride and chloroethane) are compared with ethanol for the AM1, RM1 and POA1 models. Ethanol is known to be more polar than the other substances analyzed. However, with the RM1 method this is not observed, explaining the poor results of RM1 for the dipole moment for these atoms when compared to AM1. Thus, the semi-empirical quantum chemical method POA1 was used for the COSMO calculations using MOPAC.



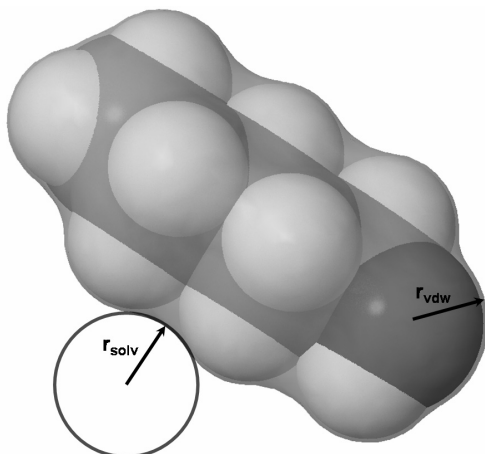


(c) POA1

**Figure 2:** Comparison of the sigma profiles of substances with Cl and ethanol for AM1, RM1 and POA1.

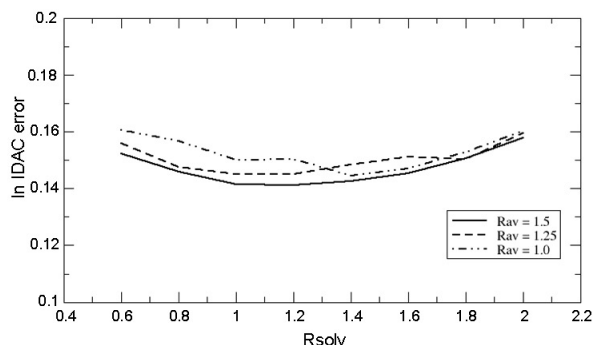
### Solvent Radius

Another important step in the COSMO calculations is the determination of the molecular cavity. For this task, the solvent radius ( $r_{\text{solv}}$ ) and atomic radii ( $r_{\text{vdw}}$ ), as shown in Figure 3, are needed. The solvent radius is used to determine the effective surface of the molecule, *i.e.* the surface accessible to the solvent. The effect of the atomic radii in the COSMO-calculations will be investigated next.

**Figure 3:** Schematic illustrating the solvent radius ( $r_{\text{solv}}$ ), the atomic radii ( $r_{\text{vdw}}$ ) and smoothed surface of the pentanol molecule.

A preliminary study, with a set of 469 experimental IDAC data points, was carried out to analyze the effect of  $r_{\text{solv}}$  and  $r_{\text{av}}$ . Figure 4 shows the average absolute error in the prediction of the logarithm of the IDAC upon varying  $r_{\text{solv}}$  and  $r_{\text{av}}$ .

For each combination of  $r_{\text{solv}}$  and  $r_{\text{av}}$  all molecules were reprocessed by MOPAC. Then, the COSMO-SAC parameters  $q$ ,  $p$ ,  $\alpha'$  and  $\beta$  were recalibrated and the average IDAC error after calibration registered. The range of mean absolute error was very small, the lowest error for  $r_{\text{av}}$  values equal to 1.5, 1.25, and 1.0 were 0.142, 0.145, and 0.145 ln-units, respectively. As can be seen in Figure 4, the model is only marginally sensitive to the solvent radius. Thus,  $r_{\text{solv}}$  was set to 1.4 for the remaining analysis,  $r_{\text{av}}$  will be further investigated in the next section.

**Figure 4:** Average absolute error in the prediction of the logarithm of the IDAC for a set of 469 data points versus  $r_{\text{solv}}$  for different  $r_{\text{av}}$ .

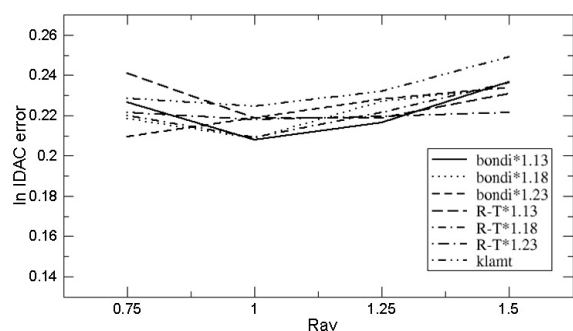
### Atomic Radii

An approximation widely used is that atoms of a given element may be regarded as hard spheres with a characteristic radius, the van der Waals radius ( $r_{\text{vdw}}$ ). In the literature, there are several tabulations of van der Waals radii, *e.g.*, Bondi (1964) and

Rowland-Taylor (1996). These radii are obtained from crystal structures.

Klamt showed that the optimal radii for COSMO calculations are around 117% of Bondi's vdWs radii, with the exception of hydrogen, for which better results were obtained with a significantly smaller value. This difference also appears in the hydrogen vdW radius of Rowland and Taylor (1.1 Å), which is shorter than Bondi's value (1.2 Å). The increase of 17% in the radii may represent an attempt to take free-volume into account.

In this work, an analysis of the effect of the atomic radii on the COSMO-SAC performance was carried out. As a basis for the atomic radii, both sets of vdW radii were considered: those of Bondi (1964) and of Rowland and Taylor (1996). For each set, different *free-volume* scaling factors were considered. In the analysis, for each pair of scaling factors and  $r_{av}$ , the parameters of the combinatorial term were re-estimated (see Eq. (3)). The energetic term parameters were recalibrated as well, but considering a larger data set with 2236 experimental IDAC. The resulting deviations are depicted in Figure 5.



**Figure 5:** Average absolute error in the prediction of the logarithm of the IDAC for a set of 2544 data points versus  $r_{av}$  for different atomic radii (Bondi, 1964; Klamt, 2000; Rowland and Taylor, 1996).

Analyzing Figure 5, we note that the prediction of the IDAC is essentially the same, with values falling in a narrow window of 0.21 to 0.23. As a result, a set of parameters with a small error but also consistent with previous values available in the literature was chosen. The parameters used are summarized in the Table 2.

### Combinatorial Term

It is well-known that the residual contribution can be orders of magnitude larger than the combinatorial one. Furthermore, the combinatorial contribution may be of the same order of magnitude as the

experimental noise for highly non-ideal mixtures. Thus, for a better fit of the combinatorial parameters, these should be estimated separately, using data with predominantly combinatorial effects. This was already accomplished in a previous work by our group (Soares, 2011).

However, when the atomic radii change, the molecular area and volume change as well. Thus, a new calibration was carried out in this work and the resulting values are also shown in Table 2. The parameter estimation was carried out using the same procedure and experimental data subset as reported by Soares (2011).

**Table 2:** Parameter set used for COSMO-SAC predictions and comparison with those originally proposed by Lin and Sandler (2002).

Quantum chemical method	This work	COSMO-SAC <sup>lit</sup>
	MOPAC semiempirical POA1	DMol3 VWN-BP/DNP
atomic radii	bondi*1.18	klamt
Solvent Radius	1.4	1.3
$q$ (Å <sup>2</sup> )	144.68	79.53
$p$	0.627	-
$\beta$	0.816	1.0
$\alpha'$ ((kcal Å <sup>4</sup> )/(e <sup>2</sup> mol))	$1.03 \times 10^4$	$1.65 \times 10^4$

## RESULTS AND DISCUSSION

In this work, the IDAC database made available in a previous work (Gerber and Soares, 2010) was expanded with data points from Thomas *et al.* (1982), Dallas and Carr (1994), Castells *et al.* (2000) and Lazzaroni *et al.* (2005). As mentioned previously, only substances known to not form hydrogen bonds were considered. The resulting database consists of 2236 binary mixtures in the temperature range of 250 K to 450 K, involving 203 different substances.

When using this database, despite our re-parameterization, COSMO-SAC still performed worse than UNIFAC (Do). Considering all data points, UNIFAC presents a mean absolute error equal to 0.12 ln-units against 0.22 from COSMO-SAC. This shows that, in general, COSMO-based models still cannot outperform group contribution methods. Apparently, simple re-calibration of the current COSMO-SAC parameters does not seem to be sufficient for a substantial improvement of the model. It might be necessary to add new parameters/formulations, or even estimate the semi-empirical QM internal parameters using IDAC data to improve the performance. Other works based on more sophisticated quantum packages (Grensemann and

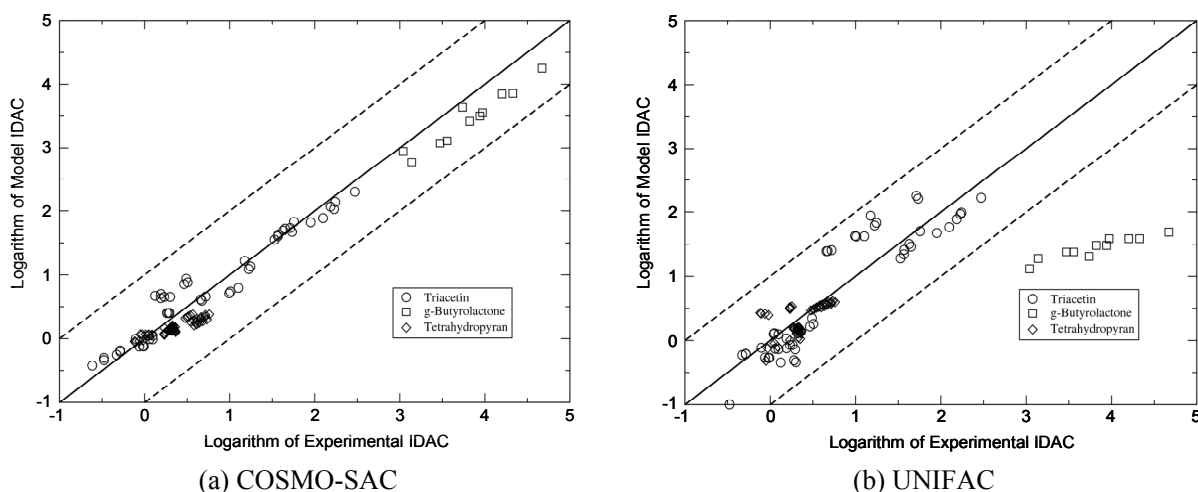
Gmehling, 2005; Wang *et al.*, 2009) also support this statement.

Nevertheless, there are situations where COSMO-SAC performs better than UNIFAC. Besides the well-known lack of differentiation between isomers, the group contribution method also fails to predict the behavior of mixtures of molecules with several functional groups or when functional groups appear in an unusual way. In order to exemplify this behavior we have selected three substances, a tri-ester (triacetin), a cyclic ester ( $\gamma$ -butyrolactone) and a cyclic ether (tetrahydropyran). The IDAC database subset for this study contains 141 points involving another 38 substances in binary mixtures. The prediction error of COSMO-SAC for this subset was 0.20, similar to

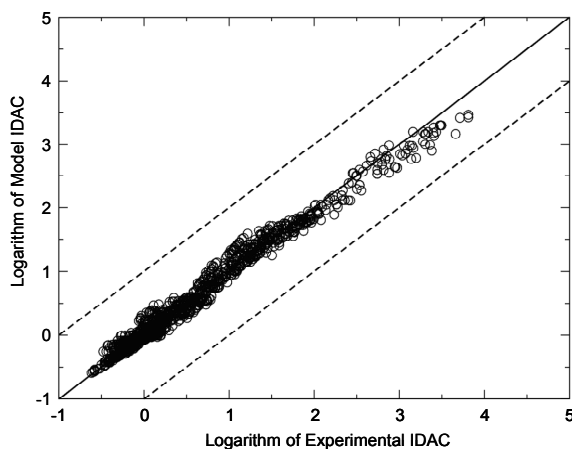
its overall error, whereas UNIFAC presents an error of 0.44, much larger than the overall one. The results for these mixtures can be seen in Figure 6.

It should be noted that, even for these complex mixtures, UNIFAC results remained in a narrow range of error (the exception was the lactone). Nonetheless, when compared to the *usual* expected prediction capacity of the model, these results are much worse.

For instance, in Figure 7, we selected the UNIFAC results for substances with at most one functional group other than a hydrocarbon portion, among them: alkanes, alkenes, amides, amines, aromatics, cycloalkanes, cycloalkenes, ethers, esters and ketones.



**Figure 6:** Prediction of the IDAC of triacetin,  $\gamma$ -butyrolactone and tetrahydropyran using COSMO-SAC (a) and UNIFAC (b).



**Figure 7:** IDAC prediction of UNIFAC for simple molecules, with at most one functional group other than a hydrocarbon portion.



From these results, it is clear that UNIFAC presents inconsistent results for molecules with several functional groups or when functional groups appear in an unusual way, especially when compared with the results for substances with at most one functional group other than a hydrocarbon portion.

Of course, the UNIFAC model performance can be improved in these cases by adding very specific and less comprehensive groups. In this regard, the UNIFAC Consortium has parameters for the *lactone* group, but these have not yet been released to the general public.

## CONCLUSIONS

In this work, the effect of several parameters of the COSMO-SAC model on its predictive power was investigated. Initially the parameters affecting the sigma profile generation were studied. They included the semi-empirical quantum chemical method and the solvent radius and the atomic radii used in the quantum chemistry package selected (MOPAC). Once these parameters were chosen, the COSMO-SAC parameters for the combinatorial term and electrostatic interactions were re-adjusted. The values of the radii do not seem to significantly affect the final prediction. Thus, values similar to those used in previous works can be chosen without adversely affecting the performance of the model.

The predictions of the calibrated model were then compared with the predictions of UNIFAC (Do). The comparisons were based on the mean absolute error for predicting the natural logarithm of the IDAC. A dataset of 2236 binary mixtures at different temperatures, involving 203 different substances, was used. It was observed that the UNIFAC model presented better overall performance with a mean absolute error of 0.12 ln-units against 0.22 from our re-calibrated COSMO-SAC model.

Even after re-calibration of the COSMO-SAC parameters, we were not able to produce a significant improvement in its prediction performance. It might be necessary to add new empirical parameters or new corrections in the model, or even estimate the semi-empirical QM internal parameters using IDAC data to improve the performance.

Upon further inspecting the results, the group contribution method had particular difficulty in cases involving molecules with several functional groups (e.g. triacetin) or when functional groups appear in an unusual way (e.g., cyclic ethers or esters such as tetrahydropyran or  $\gamma$ -butyrolactone, respectively). The IDAC database subset for this study contained

141 points involving another 38 substances in the binary mixtures. The prediction error of COSMO-SAC was 0.20, similar to its overall error, whereas UNIFAC presented an error of 0.44, much larger than its overall one.

These results indicate that COSMO-SAC provides more reliable predictions for multi-functional or more complex molecules (even with COSMO computations based on semi-empirical methods), reaffirming the future prospects of this approach.

## ACKNOWLEDGMENT

This work was partially supported by CNPq and FAPERGS-ARD2010.

## NOMENCLATURE

$A_i$	total cavity surface area of substance i	$\text{\AA}^2$
$A_i(\sigma)$	surface area of all segments with a surface charge density $\sigma$	$\text{\AA}^2$
$a_{\text{eff}}$	effective surface segment surface area	$\text{\AA}^2$
$a_{mn}$	UNIFAC group-interaction parameter between groups n and m	K
$b_{mn}$	UNIFAC group-interaction parameter between groups n and m	
$c_{\text{hb}}$	hydrogen-bonding constant	$\text{kcal } \text{\AA}^4 / \text{mol } e^2$
$c_{mn}$	UNIFAC group-interaction parameter between groups n and m	$\text{K}^{-1}$
$d_{mn}$	distance between surface segment m and n	$\text{\AA}$
$n_i(\sigma)$	number of segments with a surface charge density of $\sigma$	
$p$	empirical exponent for modified combinatorial contribution	
$p(\sigma)$	sigma profile of a mixture	
$p_i(\sigma)$	sigma profile of the pure substance i	
$q$	area normalization parameter	$\text{\AA}^2$
$q_i$	normalized surface-area parameter	
$R$	ideal gas constant	$\text{J/K mol}$
$r$	volume normalization parameter	$\text{\AA}^3$
$r_{\text{av}}$	surface-segment averaging radius	$\text{\AA}$

$r_i$	normalized volume parameter	
$r_n$	effective radius of surface segment n, assuming circular surface segments	Å
T	temperature	K
$V_i$	total cavity volume of the substance i	Å <sup>3</sup>
$x_i$	mole fraction of component i in solution	
z	coordination number	

### Greek Letters

$\alpha'$	constant for the misfit energy	kcal Å <sup>4</sup> / mol e <sup>2</sup>
$\beta$	empirical factor of residual contribution	
$\Delta G^{*res}$	free energy of surface change restoration	kcal/mol
$\Delta W(\sigma_m, \sigma_n)$	exchange energy	kcal/mol
$\Gamma(\sigma_m)$	segment activity coefficient	
$\gamma^\infty$	activity coefficient at infinite dilution	
$\gamma_i$	activity coefficient of the substance i in solution	
$\gamma_i^{comb}$	combinatorial contribution to the activity coefficient	
$\gamma_i^{res}$	residual contribution of the activity coefficient	
$\varphi_i$	normalized volume fraction	
$\Psi_{mn}$	UNIFAC group-interaction parameter between groups n and m	
$\sigma$	surface-segment charge-density distribution	e/Å <sup>2</sup>
$\sigma_{hb}$	sigma cutoff for hydrogen bonding	e/Å <sup>2</sup>
$\sigma_n^*$	surface-charge density for segment n from the COSMO output	e/Å <sup>2</sup>
$\theta_i$	normalized surface-area fraction	

### REFERENCES

Andzelm, J., Kölmel, C. and Klamt, A., Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *The Journal of Chemical Physics*, 103, (21), 9312-9321 (1995).

Baldrige, K. and Klamt, A., First principles implementation of solvent effects without outlying

charge error. *The Journal of Chemical Physics*, 106, (16), 6622-6634 (1997).

Barone, V. and Cossi, M., Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *The Journal of Physical Chemistry A*, 102, (11), 1995-2001 (1998).

Bondi, A., van der Waals Volumes and Radii. *Journal of Physical Chemistry*, 68, (3), 441-451 (1964).

Castells, C. B., Eikens, D. I. and Carr, P. W., Head-space gas chromatographic measurements of limiting activity coefficients of eleven alkanes in organic solvents at 25 °C. 1. *Journal of Chemical & Engineering Data*, 45, (2), 369-375 (2000).

Dallas, A. J. and Carr, P. W., Critical evaluation of predicted and measured gas-liquid partition coefficients in n-hexadecane. *The Journal of Physical Chemistry*, 98, (18), 4927-4939 (1994).

Dewar, M. J. S., Zebisch, E. G., Healy, E. F. and Stewart, J. J. P., Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, 107,(13), 3902-3909 (1985).

Fredenslund, A., Jones, R. L. and Prausnitz, J. M., Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal*, 21, (6), 1086-1099 (1975).

Gerber, R. P. and Soares, R. D. P., Prediction of infinite-dilution activity coefficients using UNIFAC and COSMO-SAC variants. *Industrial & Engineering Chemistry Research*, 49, (16), 7488-7496 (2010).

Gmehling, J., Li, J. and Schiller, M., A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Industrial & Engineering Chemistry Research*, 32, (1), 178-193 (1993).

Grensemann, H. and Gmehling, J., Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. *Industrial & Engineering Chemistry Research*, 44, (5), 1610-1624 (2005).

Hsieh, C.-M., Sandler, S. I. and Lin, S.-T., Improvements of COSMO-SAC for vapor-liquid and liquid-liquid equilibrium predictions. *Fluid Phase Equilibria*, 297, (1), 90-97 (2010).

Jakob, A., Grensemann, H., Lohmann, J. and Gmehling, J., Further Development of Modified UNIFAC (Dortmund): Revision and Extension 5. *Industrial & Engineering Chemistry Research*, 45, (23), 7924-7933 (2006).

Klamt, A., Conductor-like screening model for real solvents: A new approach to the quantitative

- calculation of solvation phenomena. *The Journal of Physical Chemistry*, 99, (7), 2224-2235 (1995).
- Klamt, A., COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, p. 246, Amsterdam, Ed. Elsevier (2000).
- Klamt, A., Jonas, V., Bürger, T. and Lohrenz, J. C. W., Refinement and parametrization of COSMO-RS. *The Journal of Physical Chemistry A*, 102, (26), 5074-5085 (1998).
- Lazzaroni, M. J., Bush, D., Eckert, C. A., Frank, T. C., Gupta, S. and Olson, J. D., Revision of MOSCED parameters and extension to solid solubility calculations. *Industrial & Engineering Chemistry Research*, 44, (11), 4075-4083 (2005).
- Lin, S.-T. and Sandler, S. I., A priori phase equilibrium prediction from a segment contribution solvation model. *Industrial & Engineering Chemistry Research*, 41, (5), 899-913 (2002).
- Rocha, G. B., Freire, R. O., Simas, A. M. and Stewart, J. J. P., RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *Journal of Computational Chemistry*, 27, (10), 1101-1111 (2006).
- Rowland, R. S. and Taylor, R., Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals Radii. *The Journal of Physical Chemistry*, 100, (18), 7384-7391 (1996).
- Schäfer, A., Klamt, A., Sattel, D., Lohrenz, John, C. W. and Eckert, F., COSMO implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems. *Physical Chemistry Chemical Physics*, 2, (10), 2187-2193 (2000).
- Soares, R. D. P., The Combinatorial term for COSMO-based activity coefficient models. *Industrial & Engineering Chemistry Research*, 50, (5), 3060-3063 (2011).
- Stewart, J. J. P. (n.d.), MOPAC 2009, Stewart Computational Chemistry, Version 10.251L.
- Thomas, E. R., Newman, B. A., Long, T. C., Wood, D. A. and Eckert, C. A., Limiting activity coefficients of nonpolar and polar solutes in both volatile and nonvolatile solvents by gas chromatography. *Journal of Chemical & Engineering Data*, 27, (4), 399-405 (1982).
- Truong, T. and Stefanovich, E., A new method for incorporating solvent effect into the classical, ab initio molecular orbital and density functional theory frameworks for arbitrary shape cavity. *Chemical Physics Letters*, 240, (4), 253-260 (1995).
- VRTech, [www.vrtech.com.br](http://www.vrtech.com.br) (2005).
- Wang, S., Lin, S.-T., Watanasiri, S. and Chen, C.-C., Use of GAMESS/COSMO program in support of COSMO-SAC model applications in phase equilibrium prediction calculations. *Fluid Phase Equilibria*, 276, (1), 37-45 (2009).
- Wang, S., Sandler, S. I. and Chen, C.-C., Refinement of COSMO-SAC and the applications. *Industrial & Engineering Chemistry Research*, 46, (22), 7275-7288 (2007).