Analysis of a Concept for Predicting Missing Group Interaction Parameters of the UNIFAC Model Using Connectivity Indices

André Mohs, Antje Jakob, and Jürgen Gmehling
Carl von Ossietzky University Oldenburg, Technical Chemistry, D-26111 Oldenburg, Germany

DOI 10.1002/aic.11891
Published online May 6, 2009 in Wiley InterScience (www.interscience.wiley.com).

The reliability of a method for the determination of missing group interaction parameters for the UNIFAC model[1] proposed by Gani et al.[2] was carefully examined. In this method, called CI-UNIFAC, the missing group interaction parameters are determined by means of connectivity indices. Such a method would be very desirable since it would save a lot of time and money spent for measuring VLE data to fill the gaps in the UNIFAC parameter matrix. Because of the great importance of distillation processes for the chemical industry mainly the results for vapor-liquid equilibria (VLE) were investigated using the data stored in the Dortmund Data Bank and VLE data measured. There are examples in which the method works quite well, but often the procedure leads to poor results. © 2009 American Institute of Chemical Engineers *AIChE J,* 55: 1614–1625, 2009

Keywords: connectivity index, vapor-liquid equilibria, UNIFAC, CI-UNIFAC

Introduction

Due to the fact that 60–80% of the costs of a chemical plant are caused by the separation step,[3] the reliable knowledge of the phase equilibrium behavior for the design of thermal separation processes is essential. Over the last decades, different methods have been developed for the calculation of the phase equilibrium behavior. Two approaches can be distinguished. On the one hand, there is the possibility of calculating phase equilibria by means of equations of state (EOS), and on the other hand, there are methods based on the excess Gibbs energy, like the local composition concept introduced by Wilson.[4] Both methods need only binary parameters for the calculation of the phase equilibrium behavior of multicomponent systems.

However, often the required binary data are missing. In these cases time-consuming and expensive measurements have to be carried out. Another possibility is the prediction of the missing binary data with the help of group contribution methods, such as UNIFAC[1] or modified UNIFAC.[5] These group contribution methods work with characteristic group interaction parameters from which for example all alkane-alcohol- and alcohol-alcohol-systems can be calculated using only the group interaction parameters between the alkane and the alcohol group. The great advantage is that the number of groups is much smaller than the number of compounds.

Although the number of parameters is limited, still for every group combination at least a few reliable experimental binary data for fitting the group interaction parameters are required. Another possibility would be to create artificial data for fitting the UNIFAC- or modified UNIFAC (Dortmund) model[5] parameters, e.g., by using a quantum chemical based calculation method developed by Klamt[6] and refined by Grensemann et al.[7] called COSMO-RS(Ol). However, it has been found, that in some cases the quality of the predicted artificial data is poor. A new idea to fill the gaps of the UNIFAC method was published by Gani et al.[2] who treated the parameters for the group interactions as atomic interactions and performed a multilinear regression using a

Correspondence concerning this article should be addressed to J. Gmehling at gmehling@tech.chem.uni-oldenburg.de

topological descriptor called the connectivity index. The advantage of a connectivity index based calculation of UNIFAC group interaction parameters is that group interaction parameters for unknown groups can be determined. This means that gaps in the parameter matrix can be filled easily by the use of only a small number of model specific parameters. Those parameters characterize atomic interactions, and are matched to thermodynamic data by minimizing an objective function. With those atomic interaction parameters and a molecular-specific descriptor (the connectivity index, developed by Randić[8] and refined by Keir and Hall[9]) theoretically it should be possible, to obtain the required group interaction parameters for UNIFAC by means of a multilinear regression step.

In this work, the results of the group contribution methods original UNIFAC, the modified UNIFAC (Dortmund) and also the quantum-chemical based g^E-model COSMO-RS(Ol)[7] are compared to the results of the CI-UNIFAC model where the group interaction parameters for one or more group pairs of the parameter matrix are replaced by parameters calculated by connectivity indices using the method proposed by Gani et al.[2] Furthermore VLE data was measured for systems, for which up to now gaps in the UNIFAC parameter matrix exist. The new experimental VLE data was compared with the results of the CI-method.

This article should show when the proposed procedure agrees with the experimental vapor-liquid equilibrium data and when problems occur.

Theoretical Background

The UNIFAC model

As mentioned previously, the group contribution method UNIFAC introduced by Fredenslund et al.[1] was developed to predict vapor-liquid equilibria. The required activity coefficients are calculated by a combinatorial (entropy-based) and a residual (enthalpy-based) term, which take into account the interactions between the different groups using the "solution of groups"-concept

$$\ln \gamma_i = \ln \gamma_i^{\mathrm{C}} + \ln \gamma_i^{\mathrm{R}} \tag{1}$$

The combinatorial and residual terms are calculated as

$$\ln \gamma_i^{\mathrm{C}} = 1 - V_i + \ln V_i - 5\mathrm{q}_i\left(1 - \frac{v_i}{F_{\mathrm{i}}} + \ln\frac{v_i}{F_i}\right) \tag{2}$$

$$\ln \gamma_{\mathrm{i}}^{\mathrm{R}} = \Sigma_k \mathrm{v}_k^{(i)}\left[\ln \Gamma_k - \ln \Gamma_k^{(i)}\right] \tag{3}$$

In these equations the values V_i and F_i are defined as the volume term respective surface area of the considered group divided by the mole fraction and are given as

$$\mathrm{V}_i = \frac{r_{\mathrm{i}}}{\Sigma \mathrm{x}_j \mathrm{r}_j} \tag{4}$$

$$\mathrm{F}_i = \frac{q_i}{\Sigma x_j q_j} \tag{5}$$

where q_i and r_i are calculated from relative van der Waals parameters R_K and Q_K derived from Bondi[10]

$$r_{\mathrm{i}} = \Sigma_k v_k^{(\mathrm{i})} R_k \tag{6}$$

$$q_{\mathrm{i}} = \Sigma_k v_k^{(\mathrm{i})} Q_k \tag{7}$$

In addition, the value $v_k^{(\mathrm{i})}$ in the residual term determines the number of structural groups k in molecule i. The group activity coefficient Γ_k in the mixture, and for the pure compound $\Gamma_k^{(\mathrm{i})}$ is given by

$$\ln \Gamma_k = Q_k\left[1 - \ln(\Sigma_m \Theta_m \Psi_{mk}) - \Sigma_m\left(\frac{\Theta_m \Psi_{km}}{\Sigma_n \Theta_n \Psi_{nm}}\right)\right] \tag{8}$$

where Θ_i is calculated from mole group fractions X_m and surface fractions Q_m

$$\Theta_m = \frac{Q_m X_m}{\Sigma_n Q_n X_n} \tag{9}$$

$$X_m = \frac{\Sigma_j v_m^{(j)} x_{\mathrm{j}}}{\Sigma_j \Sigma_n v_n^{(j)} x_j} \tag{10}$$

Ψ_{mn} can be calculated using the UNIFAC group interaction parameter a_{mn}, which is fitted to vapor-liquid equilibrium data

$$\Psi_{mn} = \exp\left(-\frac{a_{mn}}{T}\right) \tag{11}$$

The adjustment of the needed two parameters a_{mn} and a_{nm} is performed by minimizing an objective function. For the UNIFAC method the simplex algorithm of Nelder and Mead[11] with the following objective function is used

$$F_{\min} = \Sigma_i \Sigma_j [ln\gamma_i(UNIFAC) - ln\gamma_i(exp.)]_j^2 \tag{12}$$

The current state of the group interaction parameter matrix of the UNIFAC method is shown in Figure 1.

The CI-UNIFAC model

To extend the range of applicability of the UNIFAC model it would be desirable to predict the missing group interaction parameters without performing time-consuming vapor-liquid equilibrium measurements for systems containing the missing groups. There were several publications using topological descriptors, especially the connectivity index, for the prediction of pure component properties.[13] So, it was obvious to use those indices to predict interaction parameters for molecular fragments, since only a few characteristics of the considered fragments are needed. In 2007 Gani et al.[2] developed a method based on the well-known connectivity index proposed by Randić[8] and the modifications made by Keir and Hall.[9]

The first step to obtain the characteristic index for a certain kind of molecular fragment is to generate the hydrogen-suppressed graph of the fragment. This means that any atom in this fragment is represented with a dot (called vertex), and connections between those vertexes are shown as lines (called edges). This notation is also common in organic chemistry to illustrate different vertex degrees of alkane molecules.

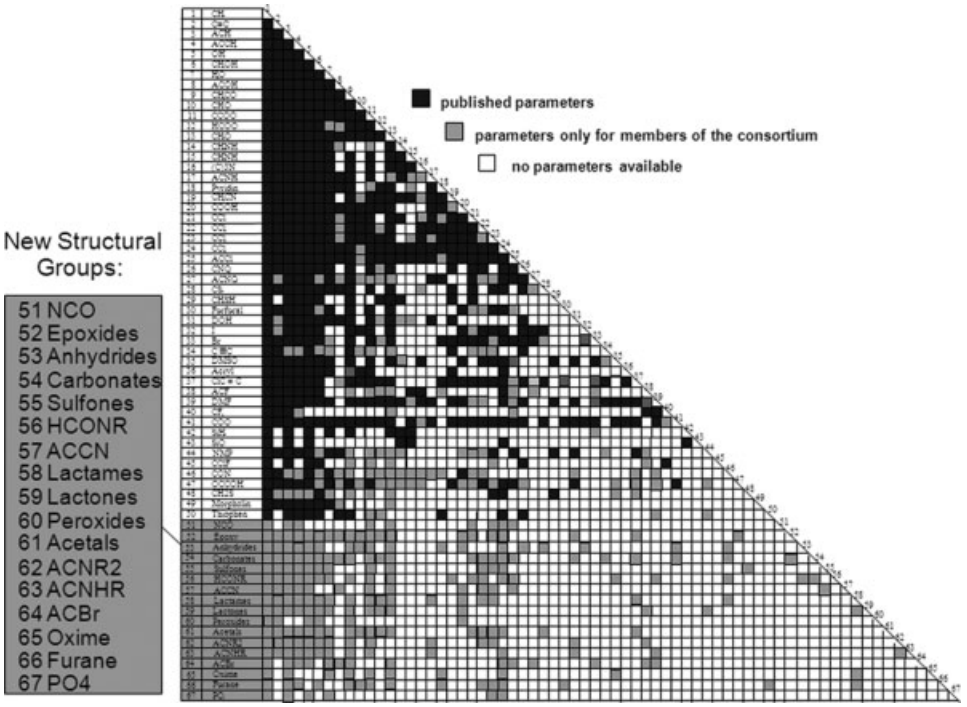

Figure 1. Current state of the UNIFAC matrix.[12]

For this graph the first calculation carried out refers to the atomic valence indices, which take into account the electronic differences between the elements of the periodic system of elements and are defined as

$$\delta^V = \frac{(Z^V - N_H)}{(Z - Z^V - 1)} \tag{13}$$

where Z is the atomic number, Z^V is the number of valence electrons, and N_H is the number of hydrogen atoms at the corresponding vertex. Thus, for every vertex in this graph an atom-characteristic index can be generated from which the connectivity indices of different orders can be obtained.

For the calculation of fragment-characteristic connectivity indices of different orders the general equation

$${}^{\mathrm{V}}\chi^{\mathrm{m}} = \Sigma_i \Pi_j \left(\frac{1}{\sqrt{\delta_j^V}} \right) \tag{14}$$

is used. For the connectivity index zeroth order, Eq. 14 becomes

$${}^{\mathrm{V}}\chi^0 = \Sigma_i \left(\frac{1}{\sqrt{\delta_i^V}} \right) \tag{15}$$

For the molecular fragment shown in Figure 2, the connectivity indices are obtained as follows:

The first step is the calculation of the atomic valence indices for each vertex

$$\delta_1 = \frac{6 - 0}{8 - 6 - 1} = 6 \tag{16}$$

$$\delta = \begin{pmatrix} 6 \\ 4 \\ 4 \end{pmatrix} \tag{17}$$

After that, the connectivity index zeroth order is calculated by summing up all reciprocal square roots of the atomic valence indices

$${}^{V}\chi^0 = \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{4}} + \frac{1}{\sqrt{4}} = 1,408 \tag{18}$$

Then the first-order connectivity index for this fragment can be calculated from the product of atomic valence indices participating at the considered bonds

$${}^{V}\chi^1 = \frac{1}{\sqrt{6*4}} + \frac{1}{\sqrt{4*4}} = 0.454 \tag{19}$$

If there are exactly two bonds between atom 1, 2 and 3 for this path a second-order connectivity index can be generated by building the product of all three atomic valence indices and summing up over all paths

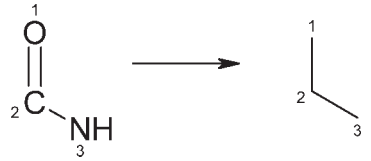

Figure 2. Example fragment for the calculation of CI's.

Table 1. Standard Atomic Interaction Parameters as well as Special Parameters for Water or Methanol Containing Systems[2]

	Normal		Water		Methanol	
	a_{mn}	a_{nm}	a_{mn}	a_{nm}	a_{mn}	a_{nm}
$\mathbf{b_{C\text{-}C}}$	977,87	−145,02	–	–	417,1	14,01
$\mathbf{b_{C\text{-}O}}$	−1134,83	401,21	−845,17	−2205,53	396,91	−94,79
$\mathbf{b_{C\text{-}N}}$	20,04	−52,78	–	–	–	–
$\mathbf{b_{O\text{-}C}}$	−11028,85	63,32	−268,83	123,53	−66,42	86,4
$\mathbf{b_{O\text{-}O}}$	11347,17	−515,78	963,82	909,49	−74,84	0,14
$\mathbf{b_{O\text{-}N}}$	0	−1,121	–	–	–	–
$\mathbf{b_{N\text{-}C}}$	45,41	194,75	–	–	–	–
$\mathbf{b_{N\text{-}O}}$	1,83	−395,38	–	–	–	–
$\mathbf{b_{N\text{-}N}}$	43,47	−4,63	–	–	–	–
$\mathbf{c_{C\text{-}C}}$	−108,11	−281,54	–	–	−22,34	−236,84
$\mathbf{c_{C\text{-}O}}$	−17,44	358,76	–	–	−29,41	−236,84
$\mathbf{c_{C\text{-}N}}$	9,94	−117,17	–	–	–	–
$\mathbf{c_{O\text{-}C}}$	−3913,23	−34	706,28	−1071,6	61,12	−97,14
$\mathbf{c_{O\text{-}O}}$	4250,92	−472,59	71,54	5821,92	93,49	−97,14
$\mathbf{c_{O\text{-}N}}$	10,82	2,27	–	–	–	–
$\mathbf{c_{N\text{-}C}}$	−32,91	450,73	–	–	–	–
$\mathbf{c_{N\text{-}O}}$	11,81	−91,91	–	–	–	–
$\mathbf{c_{N\text{-}N}}$	−38,85	−69,02	–	–	–	–
$\mathbf{d_{C\text{-}C}}$	104,61	312,4	–	–	−54,19	−250,59
$\mathbf{d_{C\text{-}O}}$	45,25	−69,53	–	–	−94,63	−250,59
$\mathbf{d_{C\text{-}N}}$	47,79	−212,21	–	–	–	–
$\mathbf{d_{O\text{-}C}}$	−633,11	195,52	–	–	−54,19	−35,2
$\mathbf{d_{O\text{-}O}}$	329,24	1026,39	–	–	−94,63	−35,2
$\mathbf{d_{O\text{-}N}}$	47,79	−3,85	–	–	–	–
$\mathbf{d_{N\text{-}C}}$	333,83	−251,6	–	–	–	–
$\mathbf{d_{N\text{-}O}}$	6,27	27,47	–	–	–	–
$\mathbf{d_{N\text{-}N}}$	327,89	−65,09	–	–	–	–
$\mathbf{e_{C\text{-}C}}$	−109,42	261,28	–	–	−3,51	–
$\mathbf{e_{C\text{-}O}}$	61,83	66,83	–	–	−14,42	–
$\mathbf{e_{C\text{-}N}}$	−6,54	20,01	–	–	–	–
$\mathbf{e_{O\text{-}C}}$	2128,63	−30,87	16,3	1,17	−3,51	–
$\mathbf{e_{O\text{-}O}}$	−2150,65	−398,08	3,92	–	−14,42	–
$\mathbf{e_{O\text{-}N}}$	−2,8	18,62	–	–	–	–
$\mathbf{e_{N\text{-}C}}$	108,41	−283,58	–	–	–	–
$\mathbf{e_{N\text{-}O}}$	30,51	−283,66	–	–	–	–
$\mathbf{e_{N\text{-}N}}$	83,23	20	–	–	–	–

$$ {}^{V}\chi^{2} = \frac{1}{\sqrt{6*4*4}} = 0.102 \tag{20} $$

So for this fragment three connectivity indices of different orders and with different content of information can be obtained.

To be able to calculate group interaction parameters not only the connectivity indices are needed, but also several other properties of the considered groups, like the number of atoms and the already mentioned atomic interaction parameters (AIPs), which are specific constants for this method. The parameter a_{mn} can then be calculated by the following equation given by Gani et al.[2]

$$
\begin{aligned}
a_{mn} = {} & \underbrace{b_{C\text{-}C}\left(A_{mn}^{CC}\right)_0 + b_{C\text{-}O}\left(A_{mn}^{CO}\right)_0 + b_{C\text{-}N}\left(A_{mn}^{CN}\right)_0 + b_{O\text{-}C}\left(A_{mn}^{OC}\right)_0 + b_{O\text{-}O}\left(A_{mn}^{OO}\right)_0 + b_{O\text{-}N}\left(A_{mn}^{ON}\right)_0 + b_{N\text{-}C}\left(A_{mn}^{NC}\right)_0 + b_{N\text{-}O}\left(A_{mn}^{NO}\right)_0 + b_{N\text{-}N}\left(A_{mn}^{NN}\right)_0}_{\text{Interactions 0th order}} \\
& + \underbrace{c_{C\text{-}C}\left(A_{mn}^{CC}\right)_1 + c_{C\text{-}O}\left(A_{mn}^{CO}\right)_1 + c_{C\text{-}N}\left(A_{mn}^{CN}\right)_1 + c_{O\text{-}C}\left(A_{mn}^{OC}\right)_1 + c_{O\text{-}O}\left(A_{mn}^{OO}\right)_1 + c_{O\text{-}N}\left(A_{mn}^{ON}\right)_1 + c_{N\text{-}C}\left(A_{mn}^{NC}\right)_1 + c_{N\text{-}O}\left(A_{mn}^{NO}\right)_1 + c_{N\text{-}N}\left(A_{mn}^{NN}\right)_1}_{\text{Interactions 1st order}} \\
& + \underbrace{d_{C\text{-}C}\left(A_{mn}^{CC}\right)_2 + d_{C\text{-}O}\left(A_{mn}^{CO}\right)_2 + d_{C\text{-}N}\left(A_{mn}^{CN}\right)_2 + d_{O\text{-}C}\left(A_{mn}^{OC}\right)_2 + d_{O\text{-}O}\left(A_{mn}^{OO}\right)_2 + d_{O\text{-}N}\left(A_{mn}^{ON}\right)_2 + d_{N\text{-}C}\left(A_{mn}^{NC}\right)_2 + d_{N\text{-}O}\left(A_{mn}^{NO}\right)_2 + d_{N\text{-}N}\left(A_{mn}^{NN}\right)_2}_{\text{Interactions 2nd order}} \\
& + \underbrace{e_{C\text{-}C}\left(A_{mn}^{CC}\right)_2 + e_{C\text{-}O}\left(A_{mn}^{CO}\right)_2 + e_{C\text{-}N}\left(A_{mn}^{CN}\right)_2 + e_{O\text{-}C}\left(A_{mn}^{OC}\right)_2 + e_{O\text{-}O}\left(A_{mn}^{OO}\right)_2 + e_{O\text{-}N}\left(A_{mn}^{ON}\right)_2 + e_{N\text{-}C}\left(A_{mn}^{NC}\right)_2 + e_{N\text{-}O}\left(A_{mn}^{NO}\right)_2 + e_{N\text{-}N}\left(A_{mn}^{NN}\right)_2}_{\text{Interactions 2nd order}}
\end{aligned}
\tag{21}
$$

Table 2. Experimental Vapor-Liquid Equilibrium Data

Nitroethane(1)/Hexanal(2) at T = 323.15 K		Hexanal(1)/Ethanediol(2) at T = 323.15 K		Ethylformate(1)/ Nitroethane(2) at T = 323.15 K		Ethylformate(1)/Ethanediol(2) at T = 298.15 K	
x_1 in mol/mol	P in kPa	x_1 in mol/mol	P in kPa	x_1 in mol/mol	P in kPa	x_1 in mol/mol	P in kPa
0,0000	5,708	0,0015	0,113	0,0000	9,672	0,0015	0,641
0,0047	5,761	0,0027	0,179	0,0016	9,818	0,0028	1,215
0,0091	5,811	0,0039	0,244	0,0032	9,955	0,0045	1,922
0,0140	5,861	0,0057	0,337	0,0069	10,275	0,0076	3,244
0,0188	5,909	0,0085	0,477	0,0105	10,587	0,0108	4,510
0,0275	5,989	0,0126	0,674	0,0140	10,887	0,0156	6,308
0,0419	6,104	0,0190	0,962	0,0203	11,419	0,0226	8,700
0,0598	6,237	0,0273	1,293	0,0291	12,156	0,0327	11,833
0,0848	6,412	0,0381	1,664	0,0420	13,244	0,0453	15,319
0,1141	6,601	0,0516	2,039	0,0583	14,624	0,0614	18,981
0,1487	6,817	0,0678	2,395	0,0786	16,315	0,0812	22,566
0,1883	7,051	0,0888	2,743	0,1032	18,362	0,1045	25,743
0,2318	7,291	0,1127	3,039	0,1319	20,728	0,1341	28,487
0,2831	7,561	0,1399	3,293	0,1674	23,623	0,1671	30,355
0,3362	7,826	0,1698	3,512	0,2062	26,770	0,2036	31,083
0,3907	8,086	0,2023	3,702	0,2483	30,139	0,2426	31,071
0,4442	8,323	0,2369	3,872	0,2922	33,633	0,2835	31,072
0,4960	8,543	0,2734	4,030	0,3375	37,198	0,3258	31,081
0,5453	8,736	0,2778	4,045	0,3833	40,770	0,3688	31,076
0,5914	8,916	0,3140	4,148	0,4287	44,311	0,3704	31,173
0,6283	9,047	0,3544	4,252	0,4290	44,296	0,4114	31,087
0,6337	9,067	0,3998	4,360	0,4716	47,579	0,4118	31,144
0,6688	9,153	0,4495	4,466	0,4732	47,697	0,4535	31,093
0,7091	9,252	0,5034	4,573	0,5160	50,972	0,4564	31,121
0,7482	9,345	0,5608	4,675	0,5168	51,012	0,5046	31,097
0,7855	9,425	0,6212	4,774	0,5645	54,608	0,5554	31,083
0,8206	9,492	0,6821	4,874	0,6135	58,289	0,6080	31,065
0,8532	9,552	0,7427	4,975	0,6631	61,996	0,6615	31,047
0,8825	9,598	0,7948	5,060	0,7123	65,667	0,7151	31,032
0,9085	9,637	0,8424	5,142	0,7604	69,258	0,7666	31,017
0,9290	9,657	0,8838	5,218	0,8056	72,645	0,8155	31,003
0,9464	9,680	0,9176	5,270	0,8476	75,794	0,8559	31,011
0,9606	9,690	0,9448	5,328	0,8815	78,354	0,8913	31,048
0,9719	9,698	0,9626	5,370	0,9109	80,573	0,9211	31,208
0,9807	9,701	0,9744	5,397	0,9353	82,447	0,9449	31,439
0,9868	9,709	0,9841	5,430	0,9546	83,912	0,9635	31,704
0,9942	9,703	0,9906	5,432	0,9696	85,066	0,9755	31,927
1,0000	9,709	1,0000	5,471	0,9792	85,814	0,9833	32,096
				0,9857	86,301	0,9890	32,249
				0,9909	86,693	0,9935	32,355
				0,9941	86,952	0,9969	32,436
				0,9973	87,181	0,9990	32,535
				0,9991	87,317	1,0000	32,560
				1,0000	87,394		

where the CI-parameters A_{mn} are calculated by group specific properties, and the multipliers symbolize the specific atomic interaction parameters which are given in Table 1. Note that the atomic interaction parameters are actually rather high, e.g., b_{O-C} or b_{O-O}

$$\left(A_{mn}^{XY}\right)_0 = \frac{n_X^{(m)V}\chi_{(n)}^0 - n_Y^{(n)V}\chi_{(m)}^0}{{}^V\chi_{(n)}^0\ {}^V\chi_{(m)}^0} = \frac{n_X^{(m)}}{{}^V\chi_{(m)}^0} - \frac{n_Y^{(n)}}{{}^V\chi_{(n)}^0} \tag{22}$$

$$\left(A_{mn}^{XY}\right)_1 = \frac{n_X^{(m)V}\chi_{(n)}^1 - n_Y^{(n)V}\chi_{(m)}^0}{{}^V\chi_{(n)}^1\ {}^V\chi_{(m)}^0} = \frac{n_X^{(m)}}{{}^V\chi_{(m)}^0} - \frac{n_Y^{(n)}}{{}^V\chi_{(n)}^1} \tag{23}$$

$$\left(A_{mn}^{XY}\right)_2 = \frac{n_X^{(m)V}\chi_{(n)}^1 - n_Y^{(n)V}\chi_{(m)}^1}{{}^V\chi_{(n)}^1\ {}^V\chi_{(m)}^1} = \frac{n_X^{(m)}}{{}^V\chi_{(m)}^1} - \frac{n_Y^{(n)}}{{}^V\chi_{(n)}^1} \tag{24}$$

$$\left(A_{mn}^{XY}\right)_3 = \frac{n_X^{(m)V}\chi_{(n)}^2 - n_Y^{(n)V}\chi_{(m)}^0}{{}^V\chi_{(n)}^2\ {}^V\chi_{(m)}^0} = \frac{n_X^{(m)}}{{}^V\chi_{(m)}^0} - \frac{n_Y^{(n)}}{{}^V\chi_{(n)}^2} \tag{25}$$

Thus, it is possible to calculate interaction parameters for the original UNIFAC model just from the structural information of the considered groups. Further developments are planned to expand the atomic interaction parameter matrix to sulfur and halides.[2] Up to now, only a maximum of three different kinds of vertex atoms (C, O and N), excluding the hydrogen is considered, because only the atomic interaction parameters between those atoms were published.[2]

Measurement of Vapor-Liquid Equilibrium Data

To be able to fit parameters for existing gaps in the UNIFAC parameter matrix, systematical measurements have to be carried out for systems containing those missing group

Table 3. CI-UNIFAC Parameters for the Considered Groups in the Deviation Calculation

UNIFAC group combinations m and n calculated by CI-UNIFAC	Parameter a_{mn}	Parameter a_{nm}
(1,3)	112,29	−46,093
(1,12)	615,29	102,355
(1/26) normal	483,42	−59,877
(1,26) modified	518,42	−57,618
(6/7) normal	−133,68	−472,69
(6/7) water	−155,315	230,84
(6/7) methanol	−421,595	159,15
(46/3)	267,50	236,375
(46/4)	1112,16	−1619,63
(46/6)	197,49	942,886
(46/8)	415,86	−2258,15
(46/9)	590,53	4435,28
(46/11)	514,82	3592,80
(46/15)	1015,95	109,11
(46/16)	−1882,7	−950,70
(46/17)	−471,05	150,435
(46/20)	308,88	12230
(46/31)	510,454	3327,89
(57/3) normal	−448,71	−333,92
(57/4) normal	199,057	−291,67
(57/5) normal	7574,95	459,73
(57/8) normal	7111,08	−2036,57
(57/13) normal	7217,96	−1024,07
(57/20) normal	7100,45	3119,10
(57/3) cyclic	1081,58	1857,37
(57/4) cyclic	1729,35	1654,49
(57/5) cyclic	−3637,67	602,64
(57/8) cyclic	120,797	2724,02
(57/13) cyclic	227,679	3736,51
(57/20) cyclic	110,165	7993,00

combinations. In this work, isothermal P-x-measurements for aliphatic nitro compounds (UNIFAC group 26), and ethylene glycol (UNIFAC group 31) have been performed with aldehydes (UNIFAC group 10), and formates (UNIFAC group 12). Those measurements were carried out with a computer driven static apparatus described in various articles.[14,15] It consists of a thermostatted measurement cell of known volume that is held at constant temperature with a thermostat by means of silicone oil. Known amounts of the purified, degassed compounds are injected in the cell using precise stepping motor driven injection pumps. The temperature is measured by a resistance thermometer with an accuracy of ±1 mK. The cell is connected to a pressure sensor which measures the equilibrium pressure of the injected mixture. The determination of the mole fraction in equilibrium is carried out starting from the feed composition using a software package connected to the Dortmund Data Bank. The measured equilibrium data are presented in Table 2.

Regarded Groups and Calculation of Interaction Parameters

Regarded groups

As mentioned previously the groups for which UNIFAC parameter pairs are calculated with the CI-UNIFAC model consist of carbon, nitrogen and oxygen. First of all it should be assumed that the parameter pair between alkanes and aromatic carbons (UNIFAC-groups 1 and 3) is missing. The

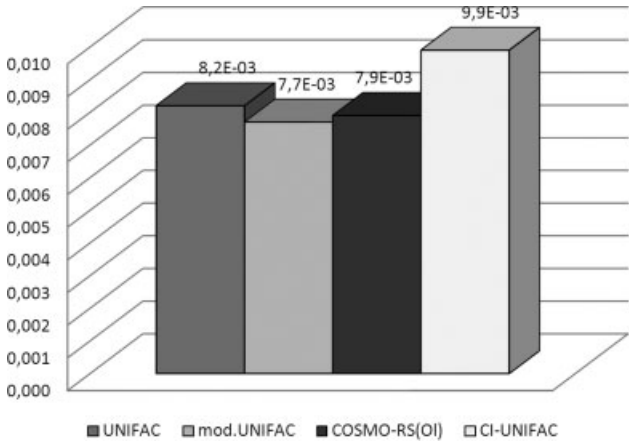

Figure 3. Mean absolute deviations in the vapor-phase mole fraction for alkanes (group 1) with aromatics (group 3).

results of the predicted VLE using the CI-method should be compared with the established methods original UNIFAC, mod. UNIFAC and COSMO-RS(Ol). After that, the accuracy of the procedure should be investigated for an increasing number of different vertex atoms. Here we will focus on the UNIFAC main group formates (UNIFAC group 12) and the nitro group (UNIFAC group 26). Then the results should be checked for the case that more parameters are missing. This means that first only one parameter pair is replaced (57/3). In the next step up to two parameter pairs are substituted by CI-UNIFAC parameters for predicting the same database containing interaction parameter pair 57/3 and up to one additional parameter pair (namely 57/4, 57/5, 57/8, 57/13, 57/20). The main group 46 (amides) will also be examined because in this group a problem concerning possible different subdivisions appears. Finally the results for the molecular groups 6 and 7 (methanol and water) are checked where the parameters can be calculated differently.

Calculation of parameter pairs

For the mentioned parameters for the group combinations (1/3) and (1/12) there exists no calculation problem with the

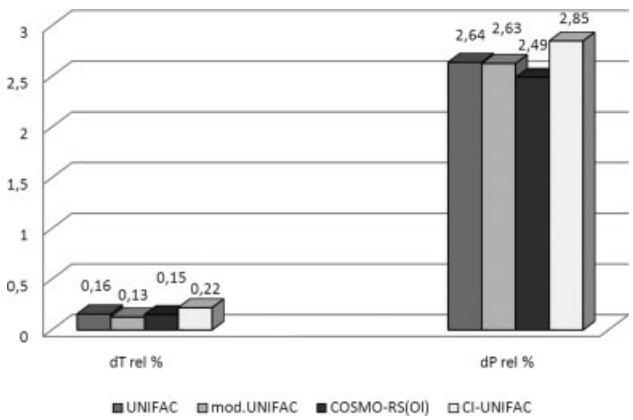

Figure 4. Mean relative deviations of T and P for alkanes (group 1) with aromatics (group 3).

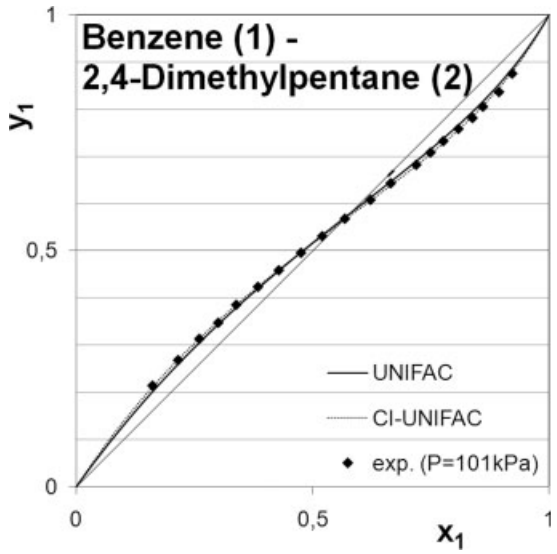

Figure 5. Exp. and predicted VLE data for the system benzene (1)-2,4-dimethylpentane at 101kPa (CI-UNIFAC, orig. UNIFAC).

introduced method. However, when going on to the nitro group (26) it has to be mentioned, that Gani et al. used a modified number of valence electrons (6 instead of 5) for nitrogen due to the problem, that in this group the nitrogen has a positive charge. With this modification[16] slightly better results for this main group are obtained. For the calculation of the cyclic vertexes in the lactame group (UNIFAC group 58) a modification for the cyclic atoms of +10 for the valence electrons was used by Gani et al. to obtain better results, due to the fact, that cyclic atoms do not behave chemically equivalent to acyclic atoms. For the group 46 a special problem occurred due to the fact, that different subdivisions are possible for this group. For the calculation of the parameters, this group was treated as O-C-N, like in the UNIFAC matrix. The calculated parameters which were obtained by a custom-built Excel routine are listed in Table 3. It can be seen, that the rather high values of the parameters for the atomic interaction cause the UNIFAC interaction parameters to be also huge, e.g., for group interaction (57/5).

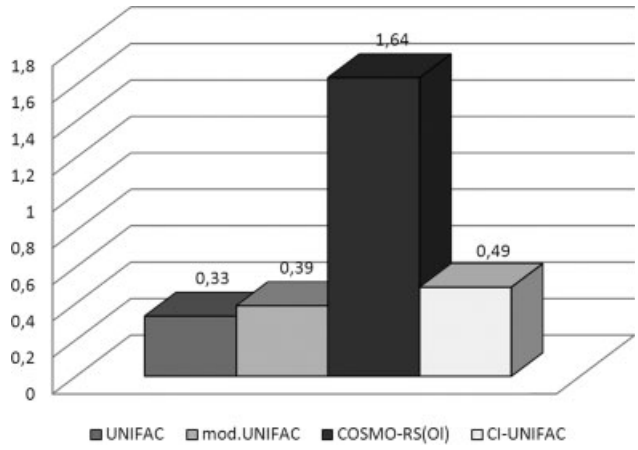

Figure 7. Mean relative deviations in T% for alkanes (group 1) with formates (group 12).

Predictions Using the Gani Model

Comparison of the Gani model with traditional g^E-models

At first a comparison is performed for alkane-aromatic systems. In this case the UNIFAC matrix element between the groups of the alkanes (1) and aromatics (3) is replaced by group interaction parameters calculated by the Gani method. Figures 3 and 4 show the calculated deviations of this procedure for 179 complete isobaric (94) and isothermal (85) consistent binary data sets, for which the thermodynamic consistency was checked via the tests of van Ness[17] and Redlich-Kister.[18] The calculations were performed with a software package implemented in the Dortmund Data Bank[19] called "MolProp" developed by Grensemann.[20] While absolute deviations for the vapor phase mole fraction y are given, relative deviations are given for the temperature and the pressure. In Figure 5 a representative y-x-diagram[21]

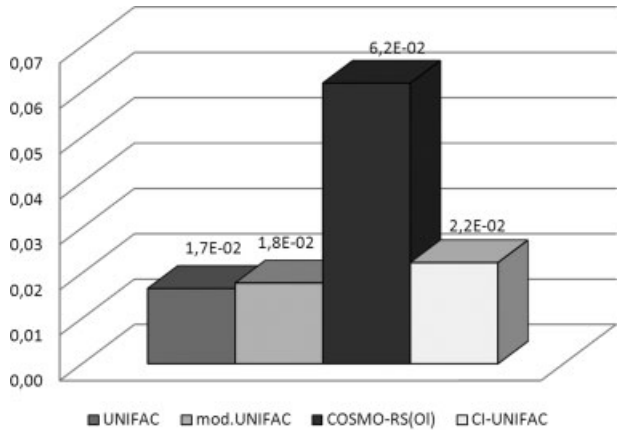

Figure 6. Mean absolute deviations in the vapor-phase mole fraction for alkanes (group 1) with formates (group 12).

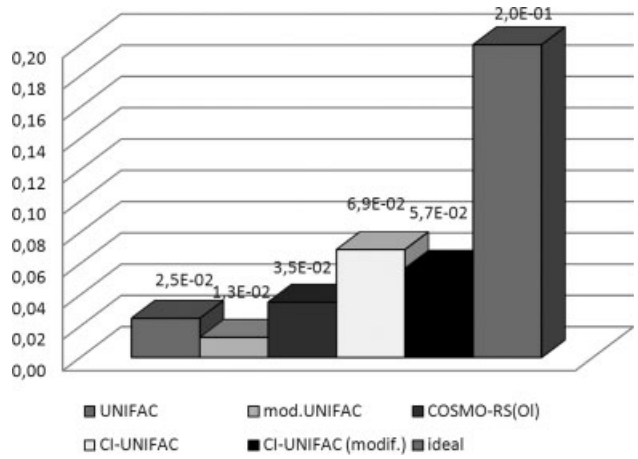

Figure 8. Mean absolute deviations in the vapor-phase mole fraction for alkanes (group 1) with aliphatic nitro compounds (group 26).

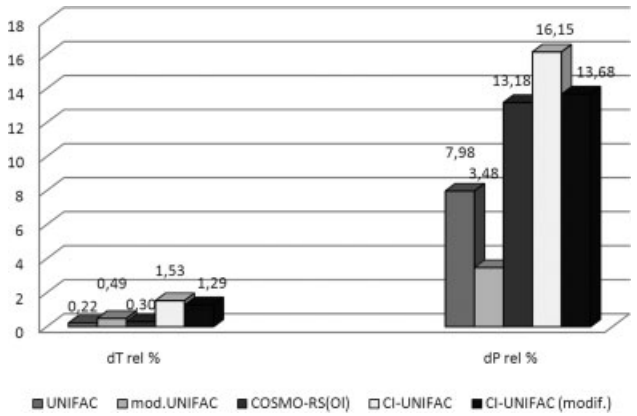

Figure 9. Mean relative deviations in T and P for alkanes (group 1) with aliphatic nitro compounds (group 26).

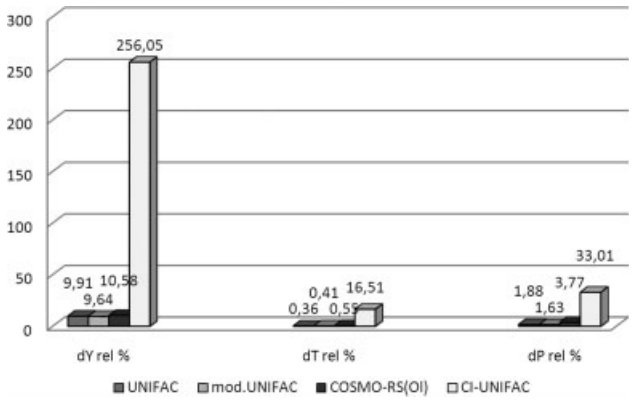

Figure 11. Deviations for the same data base as in Figure 9 but up to 2 replaced UNIFAC parameter pairs.

for benzene-2,4-dimethylpentane at atmospheric pressure is shown. From the results shown in Figures 3–5 it can be seen, that the parameters derived using the Gani method provide excellent results, although the results of the other prediction methods are still slightly better.

Analogous results can be obtained, when systems of alkanes with formates (group 1/12) are considered. In Figures 6 and the predicted deviations for five consistent isobaric VLE are shown as bar graphs. It has to be mentioned, that in this case, the CI-UNIFAC method provides even better results than the quantum chemical based model COSMO-RS(Ol). The data base used is relatively small and maybe not representative. However, still the results of original UNIFAC and mod. UNIFAC are a little better.

Regarding a group with even more hetero-atoms leads to the nitro group, represented by the UNIFAC main group 26. The deviations in the vapor-phase mole fraction, temperature and pressure for nitro compounds and alkanes are given in Figures 8 and 9. Here it is also shown that the standard calculation method (without the introduced modification for the nitrogen electrons) provides worse results as expected. Nevertheless it has to be mentioned that despite of the modification, the CI-method in this case provides poor results when compared to the other predictive models. So for example the absolute mean deviation for the CI-method is nearly factor 2.3 (5) higher than the results of UNIFAC (mod. UNIFAC).

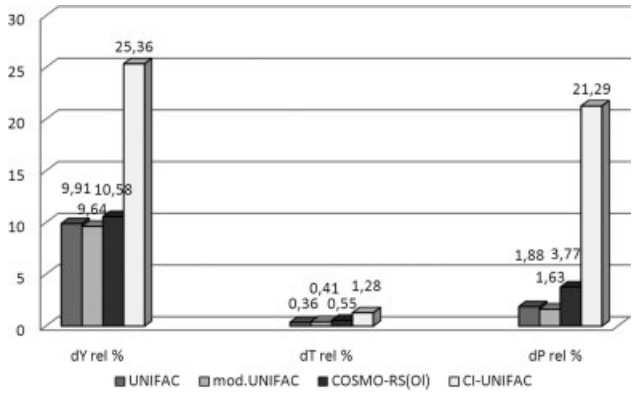

Figure 10. Mean relative deviation of the vapor-phase mole fraction, T and P for aromatics (group 3) with aromatic-cyanide compounds (group 57).

Gani et al. proposed the method for the prediction of missing UNIFAC group interaction parameters when only one parameter pair is missing. The reason for this can be seen from the results which are obtained when more parameter pairs for one main group are replaced by CI-UNIFAC parameters. In Figures 10 and 11 for the same data base containing aromatic cyanides and aromatics the results for the replacement of only the matrix element 57/3, and for the replacement of several groups with the main group 57 (3, 4, 5, 8, 13, 20) are shown. It can be seen that with the increasing number of replaced UNIFAC parameters the accuracy of the model decreases. The regarded data base in these cases consisted of 10 consistent data sets with systems containing up to two replaced UNIFAC-parameters. This time, the relative deviation in the vapor-phase mole fractions is shown.

Another problem of the model arises from the fact that different subdivisions are possible. For many functional groups there are multiple ways of subdivisions and accordingly possible different calculation results for one gap. One example is shown in Figure 12 for the main group 46, where even the number of vertex atoms can be changed which leads to different calculation possibilities. This is obviously the reason, why for group 46 (amides), rather poor results with the CI-model are obtained as can be seen from Figures 13 and 14. The data base used for the comparison consisted of 46 consistent data sets. The worst results are achieved by substituting the group 46/31 (amides/ethylene glycol) as can be recognized from the VLE results for the system N-methyl

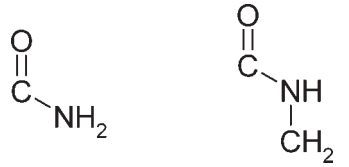

Figure 12. Example for different subdivisions in one UNIFAC group (amides).

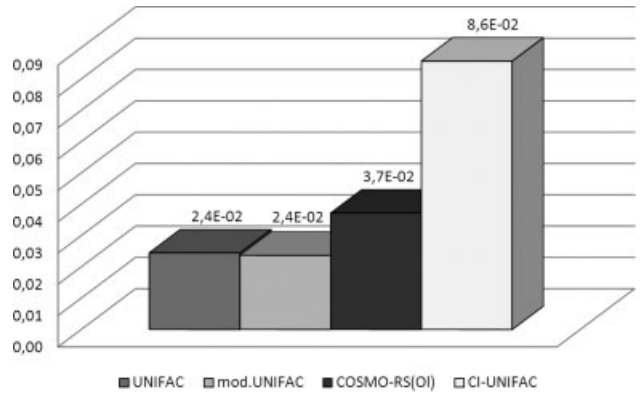

Figure 13. Mean absolute deviations in the vapor-phase mole fraction for a data base of VLE containing components with the main group 46 (amides).

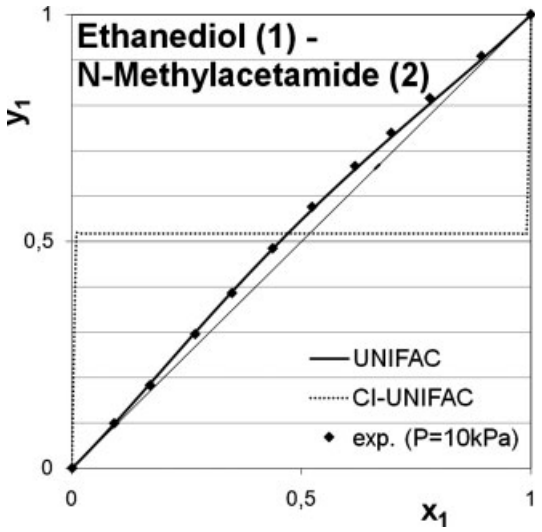

Figure 15. VLE for a system containing the UNIFAC main groups 46 and 31; CI-UNIFAC compared to orig. UNIFAC.

acetamide and ethylene glycol shown in Figure 15.[22] An explanation for this could be that the group 46 was calculated as CON like in the UNIFAC matrix. Another reason could be that ethylene glycol is in fact a molecular group and for molecular groups separate atomic interaction parameters have to be used.

For the sake of completeness another problem should be mentioned, which is caused by the different atomic interaction parameters for systems with molecular groups like water or methanol. If the system methanol-water (6/7) should be predicted, it is not clear which set of parameters should be used for the calculation. In Figure 16 the results for the different possible atomic interaction parameter sets are shown. It can be seen that reliable results are only obtained by using the water interaction parameters, so if one would like to predict this system those parameters should be used to obtain acceptable results. In Figure 17 for a representative VLE data set[23] the experimental results are shown together with the results of the three mentioned prediction possibilities. The same problem arises also when ethylene glycol is regarded as molecular group, and the interactions with water or methanol should be predicted. In this case it would make sense to determine a hierarchy which parameter sets should be used.

Comparison of the results of the Gani-model with the measured vapor-liquid equilibrium data

Finally the Gani model was used to compare the predicted results with the measured VLE data. In Figure 18, the experimental data is shown together with the predicted VLE data using the Gani method and the COSMO-RS(Ol) model. The missing interaction parameters for the CI-UNIFAC method between aldehydes and ethylene glycol (10/31), formates and ethylene glycol (12/31), as well as nitroethane with aldehydes and formates (10/26)/(12/26) were calculated by the CI-method and are listed in Table 4.

By regarding the measured data it can be recognized, that the system nitroethane–ethyl formate almost behaves ideally. The system nitroethane–hexanal shows slightly positive

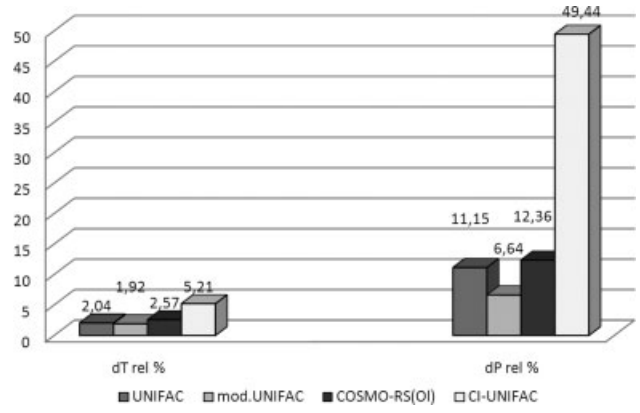

Figure 14. Mean relative deviations in T and P for the same data base containing the main group 46 (amides).

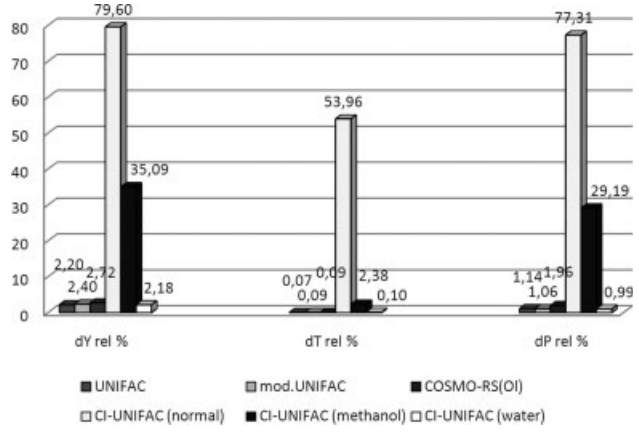

Figure 16. Mean relative deviations of the vapor-phase mole fraction, T and P for methanol (group 7) with water (group 6).

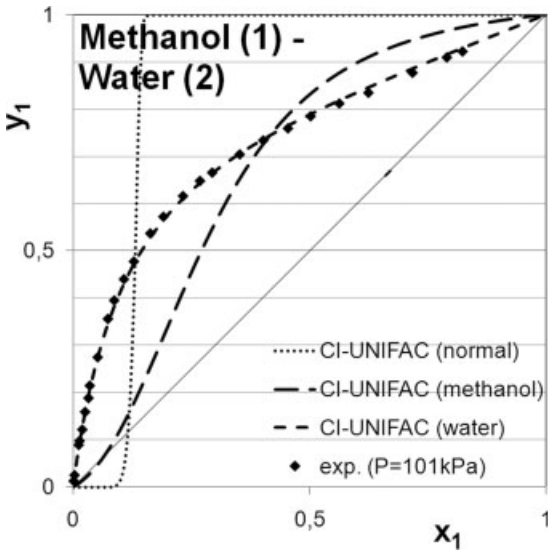

Figure 17. VLE for a system containing water and methanol; different possibilities of CI-UNIFAC prediction.

deviation from Raoults law. In comparison to that, the systems containing ethylene glycol show strong real behavior. For the system ethylene glycol/ethyl formate even a miscibility gap occurs.

From these examples it can be concluded that the method based on connectivity indices in this form cannot be used to predict the measured VLE data. However, it has to be mentioned, that the standard atomic interaction parameters have been used to calculate the glycol systems, although this is a molecular group and a special set of parameters should be used for these calculations like in water and methanol systems. If the glycol is calculated as a composition of two CH_2OH groups better results for this system can be achieved as can be seen in Figure 19 for the system ethylene glycol/ethyl formate. However, the COSMO-RS(Ol) model in most cases provides better results although the results for the glycol systems are not satisfying. Additional measurements concerning the activity coefficient at infinite dilution and the excess enthalpies are currently being carried out to fit reliable parameters for the UNIFAC- and modified UNIFAC model for the four missing group combinations.

Results and Discussion

As demonstrated, the method proposed by Gani et al. provides reliable results for common group combinations, e.g.,

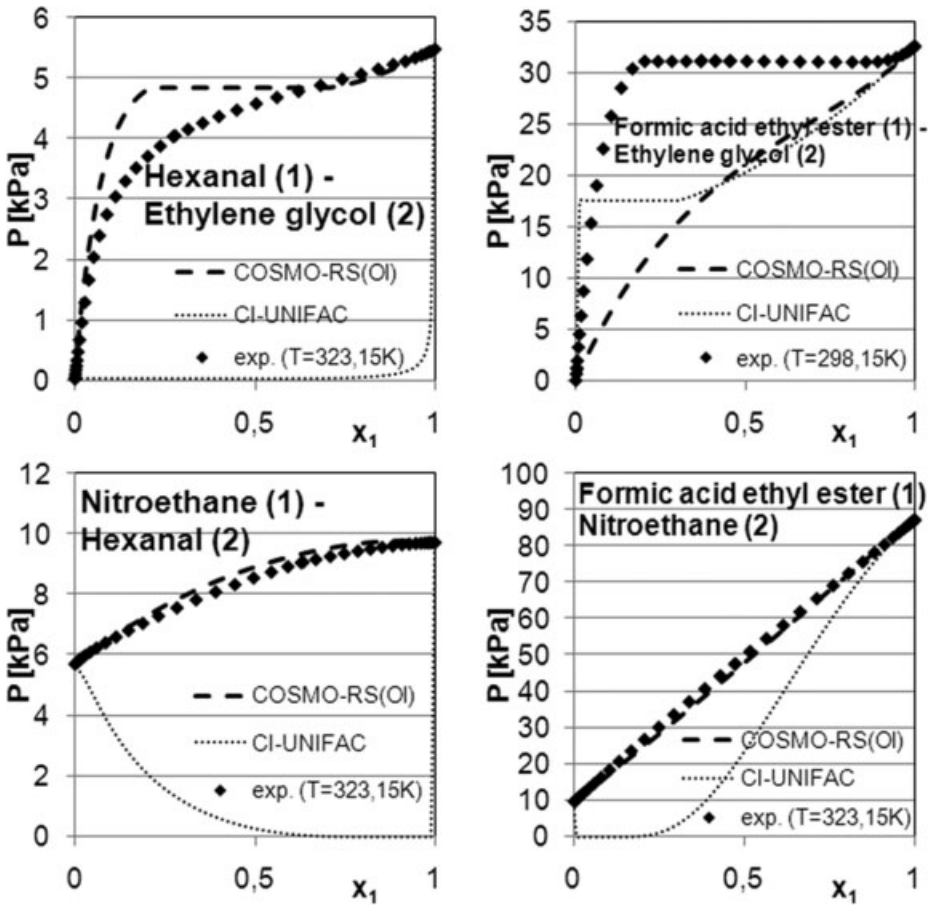

Figure 18. New VLE data of nitroethane and ethylene glycol with hexanal and ethyl formate in comparison with the predicted data obtained by using the Gani method and COSMO-RS(Ol).

aliphatics, aromatics. The reason for this could be that these groups are present in most organic molecules. In other cases, however, the procedure shows some weaknesses.

With an increasing number of hetero atoms the predicted results get worse, as shown for systems with nitro compounds. This means this procedure at least for the presented group combinations cannot be applied to fill the gaps in the UNIFAC matrix. The reason for this could be the limited number of data sets used to fit the atomic interaction parameters for the nitro group.

If more than one new group is created the numbering of the new groups plays a decisive role. This means that different results for the calculated UNIFAC interaction parameters are obtained when the numbers of the created groups are switched.

The modification introduced for cyclic atoms to calculate connectivity indices for groups like epoxides in some cases lead to results which are even worse than the prediction without the modification.

The major problem is that for one main group often multiple subdivisions are possible. This leads to different UNIFAC group interaction parameters by the calculation via connectivity indices. Speaking of chemical similarity another problem occurs during the analysis of the model, namely the over-determination of the connectivity index in some cases. Some examples are the groups 2 (C = C) and 31 (C ≡ C), groups 26 and 27 (aliphatic and aromatic nitro groups), and also groups 46 (amides) and 51 (isocyanides), for which the same parameter pairs are calculated although these groups do not behave chemically equivalent, so that different interaction parameters are required.

Conclusion

Using a comprehensive data base and new VLE data measured the results of the CI-UNIFAC model proposed by Gani et al. have been examined. While reliable results are obtained in the interpolative area of the method where the specific atomic interaction parameters have been fitted to experimental data, poor prediction results are obtained when going to group interactions which were not included in the fitting procedure. Further on in our analysis the quality of the procedure strongly decreases with the number of considered hetero atoms, but the use of a complete CI-UNIFAC matrix could give better results. In many cases, however, the method provides VLE results that are even worse than the quantum chemical approach COSMO-RS(Ol). This means, until now, it should be preferred to fill gaps in the UNIFAC matrix by using artificial data predicted using the COSMO-RS(Ol) model when no measurements are possible. Due to the fact, that COSMO-RS(Ol) sometimes predicts poor results as can be seen for example by regarding the experimental data with glycol in Figure 18, the most reliable parameters are obtained by measuring VLE data containing the missing groups and fitting the UNIFAC parameters to these data. This means that also in the future the extension of the UNIFAC matrix should be performed via reliable experimental VLE data. It has to be mentioned that Gani et al. proposed the method only to increase the applicability of the UNIFAC-model but not to improve the quality of the host model. In this article, only the parameters of the article published in 2007 have been used to calculate VLE behavior, which in the meantime, as well as some CI-calculations (e.g., the nitro group) have already been revised by the authors,[24] and could give improved results regarding VLE behavior.

Table 4. CI-UNIFAC Parameters for the Gaps in the UNIFAC Matrix that are needed to Calculate VLE

UNIFAC group combinations m and n calculated by CI-UNIFAC	Parameter a_{mn}	Parameter a_{nm}
(10,26)	−5596,13	−806,824
(12,26)	−5654,47	12968,8
(10/31)	−276,223	−2148,82
(12,31)	−229,583	2686,20

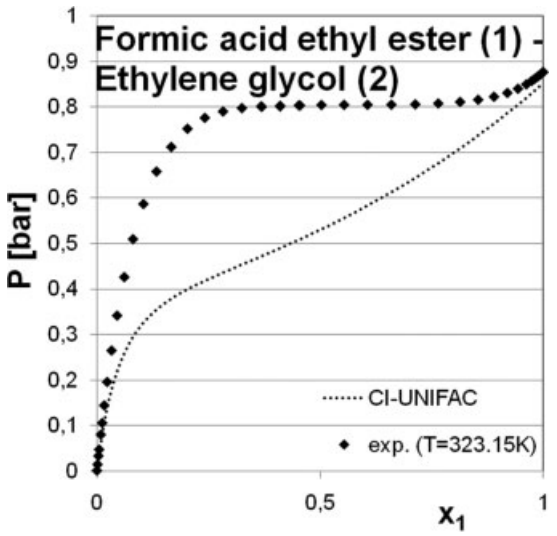

Figure 19. The same prediction as in Figure 18b when regarding ethylene glycol as structural groups (2 × CH_2OH).

Acknowledgement

The authors would like to thank DDBST GmbH for providing the Dortmund Data Bank for the examination of the model proposed by Gani et al.

Literature Cited

1. Wittig R, Lohmann J, Gmehling J. Vapor-liquid equilibria by UNIFAC group contribution. 6. revision and extension. *Ind Eng Chem Res*. 2003;42:183–188.
2. Gonzàles HE, Abildskov J, Gani R, Rousseaux P, Le Bert B. A method for prediction of UNIFAC group interaction parameters. *AlChE J*. 2007;53, 6:1620–1633.
3. Gmehling J. *UNIFAC - Ein wichtiges Werkzeug für die chemische Industrie, Software-Entwicklung in der Chemie*. Heidelberg: Springer-Verlag; 1987.
4. Wilson GM. Vapor-Liquid Equilibrium. XI. A new expression for the excess free energy of mixing. *J Am Chem Soc*. 1964;86:127–130.
5. 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.

6. Klamt A. *COSMO-RS, From quantum chemistry to fluid phase thermodynamics and drug design*. Amsterdam: Elsevier; 2005.
7. Grensemann H, Gmehling J. Performance of a conductor-like screening model for real solvents in comparison to classical group contribution methods. *Ind Eng Chem Res*. 2005;44:1610–1624.
8. Randić M. On the characterization of molecular branching. *J Am Chem Soc*. 1975;97:6609–6615.
9. Hall LH, Keir LB. Determination of topological equivalence in molecular graphs from the topological state. *Quant Struct Act Relat*. 1990;9:115–131.
10. Bondi A. *Physical properties of molecular crystals, liquids and glasses*. New York: John Wiley & Sons; 1968.
11. Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1967;7:308–313.
12. UNIFAC-consortium, http://www.unifac.org.
13. Horvath AL. Molecular design - Chemical structure generation from the properties of pure organic compounds - (series: *Studies in phyiscal and theoretical chemistry* 75). Elsevier; 1992.
14. Kolbe B, Gmehling J. Thermodynamic properties of ethanol + water. 1. vapor-liquid equilibria measurements from 90 to 150°C by the static method. *Fluid Phase Equilib*. 1985;23:213–226.
15. Rarey J, Gmehling J. Computer-operated differential static apparatus for the measurement of vapor-liquid equilibrium data. *Fluid Phase Equilib*. 1993;83:279–287.
16. Gani R, Harper PM, Hostrup M. Automatic creation of missing groups through connectivity index for pure-component property prediction. *Ind Eng Chem Res*. 2005;4:7262–7269.
17. Van Ness HC. Thermodynamics in the treatment of (vapor + liquid) equilibria. *J Chem Thermodyn*. 1995;27:113–134.
18. Redlich O, Kister A. Thermodynamics of nonelectrolyte systems (x-y-T relations in a binary system). *Ind Eng Chem*. 1948;40:341–348.
19. Dortmund Data Bank 2008, DDBST software and separation technology GmbH. www.ddbst.de.
20. Grensemann H. Oldenburg: Germany; 2004. Dissertation.
21. Richards AR, Hargreaves E. Vapor-liquid equilibria of close-boiling binary hydrocarbon mixtures. *Ind Eng Chem*. 1944;36, 9:805–808.
22. De Haan A, Fischer K, Haake M, auf der Haar O, Petri M, Gmehling J. Vapor-liquid equilibria and enthalpies of mixing for binary mixtures of N-methylacetamide with aniline, decane, ethylene glycol, naphthalene, phenol and water. *J Chem Eng Data*. 1997;42, 5:875–881.
23. Ocon J, Rebolleda F. Equilibrio Vapor-Liquido VI. Estudio del sistema binario metanol-agua. *Anal Fis Quim Ser B*. 1958;54:525–530.
24. Gonzàles HE, Abildskov J, Gani R. Computer-aided framework for pure component properties and phase equilibria prediction for organic systems. *Fluid Phase Equilib*. 2007;261:199–204.

Manuscript received Oct. 21, 2008, and revision received Feb. 6, 2009.