A Method for Prediction of UNIFAC Group Interaction Parameters

Hugo Edson Gonzàlez, Jens Abildskov, and Rafiqul Gani
CAPEC, Dept. of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

Pascal Rousseaux and Brice Le Bert
Processium, 43 Bvd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

DOI 10.1002/aic.11171

Group-contribution-based property estimation methods are suitable for obtaining quick evaluations of phase equilibrium under different conditions of temperature, pressure and composition. One of the best known and successful group-contribution (GC) based methods for prediction of liquid-phase-activity coefficients in mixtures of organic compounds is the UNIFAC method. One of the principal limitation of the UNIFAC method and all other GC based methods is that groups or group-interaction parameters needed for a specific property estimation problem may not be available. The method presented, is to be called the CI-UNIFAC (Connectivity Index – UNIFAC) method, to predict the missing UNIFAC group-interaction parameters for the calculation of vapor-liquid equilibrium (VLE). The method is described as the CI-UNIFAC for predicting missing group interaction parameters, as well as re-estimating known group interaction parameters, using a set of atom connectivities and their regressed CI interaction parameters. The performance of the CI-UNIFAC method with experimental data is compared, with a reference UNIFAC method, as well as cases where the CI-UNIFAC method is used only for the missing UNIFAC group interaction parameters.

Keywords: UNIFAC, vapor-liquid equilibrium (VLE), group-contribution methods, connectivity index (CI), interaction parameters, prediction

Introduction

The production of chemical products usually involves processing steps, such as extraction, distillation, absorption and crystallization where phase equilibria play a major role. In many chemical plants, phase equilibrium based separation processes contribute a large percentage of the total cost, especially if the separation steps involve distillation operations. Incentives to improve process efficiency in terms of energy consumption may require the examination of numerous process alternatives. Model based solution approaches where the property model is embedded within the process model offer a quick evaluation of process alternatives if the property model is able to calculate the needed properties efficiently and reliably. Therefore, it is practical to use truly predictive models of large application range, and acceptable levels of accuracy and uncertainties.

Group-contribution-based property estimation methods are very suitable for preliminary screening of process alternatives, because they are able to provide quick and reliable estimates at low-computational costs. In reverse property prediction problems, such as the design of molecules/solvents, group-contribution-based methods are ideally suited because they offer a means of generating molecular structures with

---

Correspondence concerning this article should be addressed to R. Gani at rag@kt.dtu.dk.

desired properties, as well as a predictive method for estimation of properties (Gani et al.[1]).

Among the GC based methods for prediction of liquid-phase-activity coefficients for mixtures, the UNIFAC method[2] is the best known and most successful. It has been used in many areas, for example;

(1) To calculate vapor-liquid equilibria (VLE)[2]
(2) For calculating liquid-liquid equilibria,[3]
(3) For calculating solid-liquid equilibria,[4]
(4) For determining activities in polymer solutions,[5]
(5) For determining vapor pressures of pure components,[6]
(6) For determining flash points of solvent mixtures,[7]
(7) For determining solubilities of gases,[8]
(8) For estimation of excess enthalpies.[9]

The UNIFAC method has several well-known limitations.[10] In this article, however, we are only concerned with the problem of the missing groups or group-interaction parameters needed for the calculation of liquid-phase-activity coefficients related to vapor-liquid equilibrium. The status of groups and group-interaction parameters can be seen from the published UNIFAC group-interaction parameter matrix[10,11]—it can be observed that there are still many blank spaces (indicating missing parameters) in the matrix.

If one wants to use the UNIFAC method to predict the phase equilibrium of a mixture for which interaction parameters are not available, or for which the compounds in the mixture cannot be described using the functional groups of the selected UNIFAC method, the usual procedure would be to collect the necessary experimental data, and then to regress the missing (new) interaction parameters by matching the collected data. This, however, can be time consuming, expensive, and sometimes even infeasible as it may not be possible to perform the necessary experiments. Also, since most of the available data may have already been used to estimate the UNIFAC group-interaction parameters, new parameters cannot be regressed without additional experimental data. Gmehling and coworkers have been doing this by adding new experimental data.[11]

Another way to address this problem is to create the missing groups (when needed), and to predict the missing group-interaction parameters using the available data. Connectivity indices (Kier and Hall[12]), which use information on the atoms and their connections, can be used to create new groups as shown by Gani et al.[1] for pure component property prediction. In this way, just as groups are used to represent molecules that are not included in the regression step, atom connectivities are used to represent groups that are also not included in the regression step.

In this article, a methodology is proposed to predict the missing UNIFAC group-interaction parameters for the prediction of vapor-liquid equilibrium (VLE), by using atom connectivity indices, and their interaction parameters regressed by using only the currently available experimental data. According to this methodology, when a group-interaction parameter is not available, using the (atom) connectivity indices that represent the corresponding groups and the regressed CI-interaction parameters, the missing group-interaction parameter is calculated (that is predicted) for the reference UNIFAC method (in this article, the UNIFAC method together with its groups and parameter tables reported by Hansen et al.[10] is selected as the reference UNIFAC method). The methodology, therefore, employs a parallel method for calculating the liquid-phase-activity coefficients, based on connectivity indices and their interactions, which can be used to predict the missing group interaction parameters, as well as predicting molecular interaction parameters. The article describes the developed connectivity index-based method (to be called the CI-UNIFAC method), together with a description of the parameter regression step, and an analysis of the results in different application scenarios. Note, however, that the current version is limited only to the UNIFAC groups composed of carbon, hydrogen, oxygen and/or nitrogen atoms and the "molecular" groups, water and methanol. Note also that the objective here is not to improve the accuracy of the reference UNIFAC model, but to extend its application range.

Theoretical Background

Connectivity indices

The connectivity indices ${}^{v}\chi$ are formalisms defined via graph theoretical concepts intended to describe topological characteristics of molecular structures (Kier and Hall,[12,13] Trinajstic,[14] Randic[15]). The graph theoretical treatment of the molecule starts by the construction of the hydrogen-suppressed graph of the molecular structure. For example, the differences between the molecular structure and its corresponding hydrogen-suppressed graph for acetic acid are illustrated in Figure 1. On the lefthand side of Figure 1, the representation of acetic acid with two groups is illustrated, while the corresponding molecule and group representation in terms of hydrogen suppressed graph is shown on the righthand side of Figure 1.

As shown on the righthand side of Figure 1, the nonhydrogen atoms become vertices 1, 2, 3, 4 in the graph, while the bonds become edges a, b and c. The omission of hydrogen and double bonds in the graph is compensated by the manner in which the atomic (valence) index δ^v for each vertex is defined.

The atomic index δ^v for each atom/vertex is defined as follows

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

where

Z^v is the number of valence electron in the atom, N_H is the number of hydrogen attached to the atom, Z is the atomic number.

Table 1 lists the values of atomic indices for various atoms/vertices. The defined atomic index comprises of not only information about the nature of the atom associated to

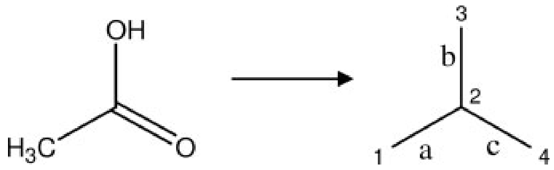

Figure 1. Molecular structure to graph.

Table 1. Atomic Indices Values for Acetic acid

Atom	1	2	3	4
Z^v	4	4	6	6
N_H	3	0	1	0
Z	6	6	8	8
δ^v	1	4	5	6

the vertex, but also about the way it is bonded to its surrounding atoms.

The zeroth-order (atomic) connectivity index ${}^v\chi^0$ is defined as a summation over the vertices of the hydrogen suppressed graph (see Figure 1).

$$ {}^v\chi^0 = \sum_i \left(\frac{1}{\sqrt{\delta_i^v}} \right) \quad i = 1, L \tag{2} $$

where L is the number of vertices (atoms) in the graph, and the δ_i^v are the valence atomic indices whose values can be obtained from Table 1 for the corresponding atom.

The first-order valence bond indices β^v can be defined for each bond, or path of length one-edge on the graph (as shown in Figure 1), by using the δ^v corresponding to the pair of bonding atoms. That is

$$ \beta^v = \delta_i^v . \delta_j^v \tag{3} $$

where i and j are the atoms connected by the bond.

Similarly, the first-order (bond) valence connectivity index ${}^v\chi^1$ is defined as a summation over the edges of the hydrogen-suppressed graph (see Figure 1)

$$ {}^v\chi^1 = \sum_i \left(\frac{1}{\sqrt{\beta_i^v}} \right) \quad i = 1, M \tag{4} $$

where M is the number of paths of length one-edge in the graph while the first-order valence bond index β^v is given by Eq. 3.

The second-order valence bond indices ε^v can be defined for each path of length two-edges on the graph (as shown in Figure 1), by using the δ^v of the corresponding atoms.

That is

$$ \varepsilon^v = \delta_i^v . \delta_j^v . \delta_k^v \tag{5} $$

where i, j and k are the atoms involved in the path.

Similarly, the second-order valence connectivity index ${}^v\chi^2$ is defined as a summation over the paths of length two-edges of the hydrogen-suppressed graph (see Figure 1)

$$ {}^v\chi^2 = \sum_i \left(\frac{1}{\sqrt{\varepsilon_i^v}} \right) \quad i = 1, N \tag{6} $$

where N is the number of paths of length two-edges in the graph while the second-order valence bond index ε^v is given by Eq. 5.

For acetic acid, whose hydrogen-suppressed graph is shown in Figure 1, and whose valence atomic index values can be found in Table 1, the zero-, first- and second-order valence connectivity indices have values of 2.355, 0.928 and 0.519, respectively.

A general formula[14] for calculating m-th order valence connectivity index is given below

$$ {}^v\chi^m = \sum_t \prod_i \left(\delta_i^v \right)^{-1/2} \quad t = 1, T; i = 1, M \tag{7} $$

where, the sum is over all type T subgraphs, with M edges using the valence δ^v values. Figure 2 shows the different type T subgraphs for each order M (with pentane as an example and $M_{\max} = 3$). Calculations of valence connectivity indices for groups are done exactly the same way as for molecules (see also Gani et al.[1]).

Valence connectivity indices of higher-orders than two can be defined to represent topological features of larger fragments. However, for the sake of simplicity and the objective of this work (predict missing group interactions), these are not employed in this article. The reason being that that the missing groups to be created through the CI-UNIFAC method will represent only a small part of the total molecular structure (none of the UNIFAC groups[10] contain more than four different atoms, not counting the molecular groups), and, therefore, higher-order terms are not necessary. Therefore, the number of needed CI-interaction parameters are kept at a manageable level. Note that the UNIFAC groups[10] are of the first-order type, and, hence, should not need higher dimensional structural parameters.

Reference UNIFAC method

The UNIFAC method, initially introduced by Fredenslund et al.,[2] is here presented in a different form that is more appropriate for purposes of computation.[16]

The UNIFAC method considers activity coefficient as composed of two additive parts – a combinatorial part to account for molecular size and shape differences, and a residual part to account for molecular interactions. For each compound i in a mixture, we have

$$ \ln \gamma_i = \ln \gamma_i^{\mathrm{COM}} + \ln \gamma_i^{\mathrm{RES}} \tag{8} $$

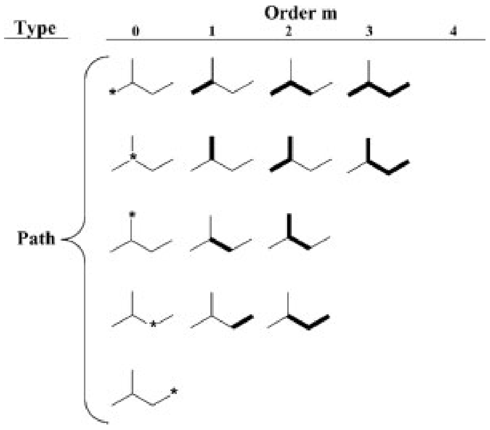

Figure 2. Types of subgraphs.

The combinatorial and residual terms, respectively, are expressed as

$$\ln \gamma_i^{\mathrm{COM}} = 1 - J_i + \ln J_i - 5q_i \left(1 - \frac{J_i}{L_i} + \ln \frac{J_i}{L_i}\right) \tag{9}$$

$$\ln \gamma_i^{\mathrm{RES}} = q_i(1 - \ln L_i) - \sum_k \left(\theta_k \frac{s_{ki}}{\eta_k} - G_{ki} \ln \frac{s_{ki}}{\eta_k}\right) \tag{10}$$

The quantities J_i and L_i are given by the following relations

$$J_i = \frac{r_i}{\sum_j r_j x_j} \tag{11}$$

$$L_i = \frac{q_i}{\sum_j q_j r_j} \tag{12}$$

In addition, the following definitions apply

$$r_i = \sum_k v_k^{(i)} R_k \tag{13}$$

$$q_i = \sum_k v_k^{(i)} Q_k \tag{14}$$

$$G_k = v_k^{(i)} Q_k \tag{15}$$

$$\theta_k = \sum_i G_{ki} x_i \tag{16}$$

$$s_{ni} = \sum_m G_{mi} \tau_{mn} \tag{17}$$

$$\eta_n = \sum_i s_{ni} x_i \tag{18}$$

and

$$\tau_{mn} = \exp \frac{-a_{mn}}{T} \tag{19}$$

Subscript i identifies species, and j is a dummy index running over all species. Subscript k identifies subgroups, subscript n identifies main groups, and m is a dummy index running over all main groups. The quantity $v_k^{(i)}$ is the number of subgroups of type k in a molecule of species i. Values of the parameters R_k and Q_k are usually calculated by the rules of Bondi,[17] while the group-interaction parameters a_{mn} are regressed by matching phase equilibrium data (the published set of Hansen et al.[10] is used in this article).

To use the reference UNIFAC method for predicting VLE of a mixture, one needs to give the molecular concentrations x_i for each compound, the group surface and volume parameters Q_k and R_k, and the group-interaction parameters a_{mk} for all the groups present in the mixture. If a new group is necessary to represent part of a molecular structure, it will need to be created (Gani et al.[1]), and the corresponding group parameters estimated. If a group already exists, but not the group interaction parameter, then only the necessary group interaction parameters will need to be estimated.

Development of the CI-UNIFAC method

In order to be able to estimate the missing UNIFAC group-interaction parameters for UNIFAC groups containing C, H, O and/or N atoms, the following procedure is applied:

(1) A relation is derived between the group-interaction parameters a_{mn} and:

- The number of C-atoms in each group,
- The number of O-atoms in each group,
- The number of N-atoms in each group,
- The connectivity indices for each group,
- CI-interaction parameters.

(2) Existing VLE data is used to regress the CI-interaction parameters, while the atom numbers and the connectivity indices are obtained directly from the group definition.

(3) Using the regressed CI-interaction parameters and the derived relation, the missing group-interaction parameters a_{mn} are predicted.

Relation between group- and CI-interaction parameters

The CI interactions are considered at 4 different levels:

- Level 1: interaction between 0th and 0th order connectivity indices
- Level 2: interaction between 0th and 1st order connectivity indices
- Level 3: interaction between 1st and 1st order connectivity indices
- Level 4: interaction between 0th and 2nd order connectivity indices

Note that the interaction between 1st and 2nd, 2nd and 2nd orders are neglected because of the relatively small size and structure of the UNIFAC groups (compared to the molecule they represent). For two groups represented by m and n, where m is less than n, the following two relations apply (counting all interactions at all levels):

$$\begin{aligned}
a_{mn} = {} & \underbrace{b_{\mathrm{C-C}}(A_{mn}^{\mathrm{CC}})_0 + b_{\mathrm{C-O}}(A_{mn}^{\mathrm{CO}})_0 + b_{\mathrm{C-N}}(A_{mn}^{\mathrm{CN}})_0 + b_{\mathrm{O-C}}(A_{mn}^{\mathrm{OC}})_0 + b_{\mathrm{O-O}}(A_{mn}^{\mathrm{OO}})_0 + b_{\mathrm{O-N}}(A_{mn}^{\mathrm{ON}})_0 + b_{\mathrm{N-C}}(A_{mn}^{\mathrm{NC}})_0 + b_{\mathrm{N-O}}(A_{mn}^{\mathrm{NO}})_0 + b_{\mathrm{N-N}}(A_{mn}^{\mathrm{NN}})_0}_{0^{\mathrm{th}}-\text{order interactions}} \\
& + \underbrace{c_{\mathrm{C-C}}(A_{mn}^{\mathrm{CC}})_1 + c_{\mathrm{C-O}}(A_{mn}^{\mathrm{CO}})_1 + c_{\mathrm{C-N}}(A_{mn}^{\mathrm{CN}})_1 + c_{\mathrm{O-C}}(A_{mn}^{\mathrm{OC}})_1 + c_{\mathrm{O-O}}(A_{mn}^{\mathrm{OO}})_1 + c_{\mathrm{O-N}}(A_{mn}^{\mathrm{ON}})_1 + c_{\mathrm{N-C}}(A_{mn}^{\mathrm{NC}})_1 + c_{\mathrm{N-O}}(A_{mn}^{\mathrm{NO}})_1 + c_{\mathrm{N-N}}(A_{mn}^{\mathrm{NN}})_1}_{1^{\mathrm{st}}-\text{order interactions}} \\
& + \underbrace{d_{\mathrm{C-C}}(A_{mn}^{\mathrm{CC}})_2 + d_{\mathrm{C-O}}(A_{mn}^{\mathrm{CO}})_2 + d_{\mathrm{C-N}}(A_{mn}^{\mathrm{CN}})_2 + d_{\mathrm{O-C}}(A_{mn}^{\mathrm{OC}})_2 + d_{\mathrm{O-O}}(A_{mn}^{\mathrm{OO}})_2 + d_{\mathrm{O-N}}(A_{mn}^{\mathrm{ON}})_2 + d_{\mathrm{N-C}}(A_{mn}^{\mathrm{NC}})_2 + d_{\mathrm{N-O}}(A_{mn}^{\mathrm{NO}})_2 + d_{\mathrm{N-N}}(A_{mn}^{\mathrm{NN}})_2}_{2^{\mathrm{nd}}-\text{order interactions}} \\
& + \underbrace{e_{\mathrm{C-C}}(A_{mn}^{\mathrm{CC}})_3 + e_{\mathrm{C-O}}(A_{mn}^{\mathrm{CO}})_3 + e_{\mathrm{C-N}}(A_{mn}^{\mathrm{CN}})_3 + e_{\mathrm{O-C}}(A_{mn}^{\mathrm{OC}})_3 + e_{\mathrm{O-O}}(A_{mn}^{\mathrm{OO}})_3 + e_{\mathrm{O-N}}(A_{mn}^{\mathrm{ON}})_3 + e_{\mathrm{N-C}}(A_{mn}^{\mathrm{NC}})_3 + e_{\mathrm{N-O}}(A_{mn}^{\mathrm{NO}})_3 + e_{\mathrm{N-N}}(A_{mn}^{\mathrm{NN}})_3}_{3^{\mathrm{rd}}-\text{order interactions}}
\end{aligned} \tag{20}$$

$$
\begin{aligned}
a_{nm} = & \underbrace{\overline{b_{\mathrm{C-C}}}(A^{\mathrm{CC}}_{nm})_0 + \overline{b_{\mathrm{C-O}}}(A^{\mathrm{CO}}_{nm})_0 + \overline{b_{\mathrm{C-N}}}(A^{\mathrm{CN}}_{nm})_0 + \overline{b_{\mathrm{O-C}}}(A^{\mathrm{OC}}_{nm})_0 + \overline{b_{\mathrm{O-O}}}(A^{\mathrm{OO}}_{nm})_0 + \overline{b_{\mathrm{O-N}}}(A^{\mathrm{ON}}_{nm})_0 + \overline{b_{\mathrm{N-C}}}(A^{\mathrm{NC}}_{nm})_0 + \overline{b_{\mathrm{N-O}}}(A^{\mathrm{NO}}_{nm})_0 + \overline{b_{\mathrm{N-N}}}(A^{\mathrm{NN}}_{nm})_0}_{0^{\mathrm{th}}-\text{order interactions}} \\
& + \underbrace{\overline{c_{\mathrm{C-C}}}(A^{\mathrm{CC}}_{nm})_1 + \overline{c_{\mathrm{C-O}}}(A^{\mathrm{CO}}_{nm})_1 + \overline{c_{\mathrm{C-N}}}(A^{\mathrm{CN}}_{nm})_1 + \overline{c_{\mathrm{O-C}}}(A^{\mathrm{OC}}_{nm})_1 + \overline{c_{\mathrm{O-O}}}(A^{\mathrm{OO}}_{nm})_1 + \overline{c_{\mathrm{O-N}}}(A^{\mathrm{ON}}_{nm})_1 + \overline{c_{\mathrm{N-C}}}(A^{\mathrm{NC}}_{nm})_1 + \overline{c_{\mathrm{N-O}}}(A^{\mathrm{NO}}_{mn})_1 + \overline{c_{\mathrm{N-N}}}(A^{\mathrm{NN}}_{nm})_1}_{1^{\mathrm{st}}-\text{order interactions}} \\
& + \underbrace{\overline{d_{\mathrm{C-C}}}(A^{\mathrm{CC}}_{nm})_2 + \overline{d_{\mathrm{C-O}}}(A^{\mathrm{CO}}_{nm})_2 + \overline{d_{\mathrm{C-N}}}(A^{\mathrm{CN}}_{nm})_2 + \overline{d_{\mathrm{O-C}}}(A^{\mathrm{OC}}_{nm})_2 + \overline{d_{\mathrm{O-O}}}(A^{\mathrm{OO}}_{nm})_2 + \overline{d_{\mathrm{O-N}}}(A^{\mathrm{ON}}_{nm})_2 + \overline{d_{\mathrm{N-C}}}(A^{\mathrm{NC}}_{nm})_2 + \overline{d_{\mathrm{N-O}}}(A^{\mathrm{NO}}_{nm})_2 + \overline{d_{\mathrm{N-N}}}(A^{\mathrm{NN}}_{nm})_2}_{2^{\mathrm{nd}}-\text{order interactions}} \\
& + \underbrace{\overline{e_{\mathrm{C-C}}}(A^{\mathrm{CC}}_{nm})_3 + \overline{e_{\mathrm{C-O}}}(A^{\mathrm{CO}}_{nm})_3 + \overline{e_{\mathrm{C-N}}}(A^{\mathrm{CN}}_{nm})_3 + \overline{e_{\mathrm{O-C}}}(A^{\mathrm{OC}}_{nm})_3 + \overline{e_{\mathrm{O-O}}}(A^{\mathrm{OO}}_{nm})_3 + \overline{e_{\mathrm{O-N}}}(A^{\mathrm{ON}}_{nm})_3 + \overline{e_{\mathrm{N-C}}}(A^{\mathrm{NC}}_{nm})_3 + \overline{e_{\mathrm{N-O}}}(A^{\mathrm{NO}}_{nm})_3 + \overline{e_{\mathrm{N-N}}}(A^{\mathrm{NN}}_{nm})_3}_{3^{\mathrm{rd}}-\text{order interactions}} \qquad (21)
\end{aligned}
$$

In Eqs. 20–21, b_{X-Y}, c_{X-Y}, d_{X-Y}, e_{X-Y}, $\overline{b_{X-Y}}$, $\overline{c_{X-Y}}$, $\overline{d_{X-Y}}$, $\overline{e_{X-Y}}$ are the interaction parameters (whose values are obtained through regression of VLE data) between atoms of type X, and atoms of type Y (where X and Y represent C, O and N atoms on a H-atom free basis); $(A^{XY}_{mn})_i$ are the coefficients to be used with the corresponding CI-interactions, that is, at the order i, between atoms of type X, and atoms of type Y, in the interaction between group m and group n, and which are defined by the following equations for each order

$$\left(A^{XY}_{mn}\right)_0 = \frac{n^{(m)}_X\,{}^v\chi^0_{(n)} - n^{(n)}_Y\,{}^v\chi^0_{(m)}}{{}^v\chi^0_{(n)}\,{}^v\chi^0_{(m)}} \qquad (22)$$

$$\left(A^{XY}_{mn}\right)_1 = \frac{n^{(m)}_X\,{}^v\chi^1_{(n)} - n^{(n)}_Y\,{}^v\chi^0_{(m)}}{{}^v\chi^1_{(n)}\,{}^v\chi^0_{(m)}} \qquad (23)$$

$$\left(A^{XY}_{mn}\right)_2 = \frac{n^{(m)v}_X\chi^1_{(n)} - n^{(n)v}_Y\chi^1_{(m)}}{{}^v\chi^1_{(n)}\,{}^v\chi^1_{(m)}} \qquad (24)$$

$$\left(A^{XY}_{mn}\right)_3 = \frac{n^{(m)}_X\,{}^v\chi^2_{(n)} - n^{(n)}_Y\,{}^v\chi^0_{(m)}}{{}^v\chi^2_{(n)}\,{}^v\chi^0_{(m)}} \qquad (25)$$

where $n^{(m)}_X$ is the number of atoms of type X in the group m; ${}^v\chi^i{}_{(m)}$ is the i-th order valence connectivity index for the group m.

Note that Eqs. 20–25 can, in principle, be written for any sets of groups containing three atoms (not counting the H atom). A check of the UNIFAC groups,[10] show that none of the 50 UNIFAC (main) groups contain more than three atoms (not counting the H atom and molecular groups). Therefore, addition of S and halide atoms will not increase the size of Eqs. 20–21, as we would still consider only a maximum of three atoms (besides the H atom). Based on the available UNIFAC groups,[10] for S atoms, only the interactions between C and S need to be considered, while for Cl and F atoms, only the interactions between C, Cl and F need to be considered. Therefore, the number of CI-interaction parameters remains at a manageable level.

Note that the actual number of CI-interaction parameters needed for any missing UNIFAC group interaction parameters are quite small since not all the terms in Eqs. 20–21 are needed (as highlighted through Eqs. 30–31 in case study 1).

Table 2 lists the UNIFAC groups[10] containing only C, H, O and N atoms and molecular groups water and methanol, along with their index numbers, atom stoichiometry and 0th, 1st, and 2nd-order connectivity indices. As the group ACH is used to represent aromatic rings, its 0th and 1st-order connectivity indices are calculated assuming that each aromatic carbon is attached to two similar groups, and that these three atoms/vertices are included in the same group. This is not the case for the second-order connectivity index. Note also that for this first version of the CI-UNIFAC method, the groups COO and CON, which are available in Hansen et al.[10] have not been considered, but will be added in the next version of the parameter tables.

To cover the 18 UNIFAC (main) groups containing only C, H, O and N atoms, which we call "functional groups", that is, excluding the "molecular" groups, water and methanol, represented in Table 2, 32 + 40 CI-interaction parameters are needed, instead of 18*17 = 306 group-interaction parameters in the reference UNIFAC method (including water and methanol interaction parameters, this number is 382). Note that 32 parameters are needed for groups with C, H and O atoms (or for groups with a maximum of three atoms), while an additional 40 parameters are needed when groups with N atoms are added (or for groups with a maximum of four atoms). This means that for future additions of atoms, if they do not have more than four atoms (UNIFAC groups do not have more than four atoms, except for molecular groups), only 40 additional parameters will be needed for each atom. However, for the addition of S atom groups, the UNIFAC group table[10] has only the (main) groups containing C, H

Table 2. Groups, Stoichiometry and Connectivities

Index	Group	n_C	n_O	n_N	χ_0	χ_1	χ_2
1	CH_2	1	0	0	0.70711	/	/
2	C=C	2	0	0	1.000	0.250	/
3	ACH	1	0	0	0.57735	0.66667	0.19245
4	$ACCH_2$	2	0	0	1.20711	0.35355	/
5	OH	0	1	0	0.44721	/	/
6	CH_3OH	1	1	0	1.44721	0.44721	/
7	H_2O	0	1	0	0.500	/	/
8	ACOH	1	1	0	0.94721	0.22361	/
9	CH_2CO	2	1	0	1.61535	0.55768	0.14434
10	COH	1	1	0	0.98560	0.23570	/
11	CCOO	2	2	0	1.81650	0.65825	0.28746
12	HCOO	1	2	0	1.39385	0.47140	0.09622
13	CH_2O	1	1	0	1.11535	0.28867	/
14	CNH_2	1	0	1	1.07735	0.28867	/
15	CNH	1	0	1	1.000	0.250	/
16	$(C)_3N$	3	0	1	1.94721	0.72361	0.33541
17	$ACNH_2$	1	0	1	1.07735	0.28867	/
19	CCN	2	0	1	1.44721	0.47361	0.11180
20	COOH	1	2	0	1.35546	0.42773	0.09129
26	CNO_2	1	2	1	1.72474	0.53746	0.23471

Table 3. Functional Groups CI-Interaction Parameters

CI-interaction Parameter (i-j)	Value	CI-interaction Parameter (j-i)	Value
b_{C-C}	977.87	$\overline{b_{C-C}}$	−145.02
c_{C-C}	−108.11	$\overline{c_{C-C}}$	−281.54
d_{C-C}	104.61	$\overline{d_{C-C}}$	321.40
e_{C-C}	−109.42	$\overline{e_{C-C}}$	261.28
b_{C-O}	−1134.83	$\overline{b_{C-O}}$	401.21
c_{C-O}	−17.44	$\overline{c_{C-O}}$	358.76
d_{C-O}	45.25	$\overline{d_{C-O}}$	−69.53
b_{C-N}	20.04	$\overline{b_{C-N}}$	−52.78
c_{C-N}	9.94	$\overline{c_{C-N}}$	−117.17
d_{C-N}	47.79	$\overline{d_{C-N}}$	−212.21
e_{C-N}	−6.54	$\overline{e_{C-N}}$	20.01
e_{C-O}	61.83	$\overline{e_{C-O}}$	66.83
b_{O-C}	−11028.85	$\overline{b_{O-C}}$	63.32
c_{O-C}	−3913.23	$\overline{c_{O-C}}$	−34.00
d_{O-C}	−633.11	$\overline{d_{O-C}}$	195.52
e_{O-C}	2128.63	$\overline{e_{O-C}}$	−30.87
b_{O-O}	11347.17	$\overline{b_{O-O}}$	−515.78
c_{O-O}	4250.92	$\overline{c_{O-O}}$	−472.59
d_{O-O}	329.24	$\overline{d_{O-O}}$	1026.39
e_{O-O}	−2150.65	$\overline{e_{O-O}}$	−398.08
b_{O-N}	0.00	$\overline{b_{O-N}}$	−1.121
c_{O-N}	10.82	$\overline{c_{O-N}}$	2.27
d_{O-N}	47.79	$\overline{d_{O-N}}$	−3.85
e_{O-N}	−2.80	$\overline{e_{O-N}}$	18.62
b_{N-C}	45.41	$\overline{b_{N-C}}$	194.75
c_{N-C}	−32.91	$\overline{c_{N-C}}$	450.73
d_{N-C}	333.83	$\overline{d_{N-C}}$	−251.60
e_{N-C}	108.41	$\overline{e_{N-C}}$	−283.58
b_{N-O}	1.83	$\overline{b_{N-O}}$	−395.38
c_{N-O}	11.81	$\overline{c_{N-O}}$	−91.91
d_{N-O}	6.27	$\overline{d_{N-O}}$	27.47
e_{N-O}	30.51	$\overline{e_{N-O}}$	−283.66
b_{N-N}	43.47	$\overline{b_{N-N}}$	−4.63
c_{N-N}	−38.85	$\overline{c_{N-N}}$	−69.02
d_{N-N}	327.89	$\overline{d_{N-N}}$	−65.09
e_{N-N}	83.23	$\overline{e_{N-N}}$	20.00

and S atoms, meaning that only interactions between C-S and S-S connectivities would be needed (that is 24 additional parameters). If groups with C, H, Cl and F atoms are considered, 64 additional CI-interaction parameters would be needed.

A set of 32 CI-interaction parameters is used to calculate methanol interactions (only for C, H and O atoms), and another 32 CI-interaction parameters for water interactions (only for C, H and O atoms). Therefore, a total of 72 + 32 + 32 = 136 connectivity index parameters are needed to estimate all of the UNIFAC group-interaction parameters for groups containing only C, H, O and N atoms plus water and methanol interactions (with groups containing C, H and O atom groups). This means that less than 50% parameters are needed for the CI-UNIFAC method to cover the interaction parameters for the 20 groups listed in Table 2.

Parameter Optimization

Values for the set of CI-interaction parameters (see Eqs. 20–21) have been determined by regression using a database involving 265 binary data sets (4,408 data-points) representing vapor-liquid equilibrium for the selected 18 functional groups, 36 data sets (636 data-points) for methanol, and 40 data sets (515 data-points) for water. The parameter estimations of this work are exclusively, based only on isothermal data. Isothermal data has been used, because in this case, it is not necessary to specify the vapor pressure or its dependence upon temperature.[18] The plan has been to use other data for purposes of validation. The VLE data sets were extracted from the CAPEC Database[19] and Processium's online database e-thermo®,[20] and were checked for thermodynamic consistency using the test of Van Ness.[21]

Due to the differences in the units and the magnitude of pressure values in the isothermal data sets, the following objective function was chosen, which is expressed as the average relative pressure quadratic deviation

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{\exp} - P_{\mathrm{calc}}}{P_{\exp}}\right)^2 \tag{26}$$

where N is the number of experimental data points used in the estimation.

For systems not needing association terms, the equilibrium pressure was calculated as follows

$$P_{\mathrm{calc}} = \sum_i x_i \gamma_i P_i^{\mathrm{sat}}(POY_i) \tag{27}$$

where i is an index running over all species in the mixture, and POY_i is the Poynting correction.

For systems needing an association term, chemical equilibrium were solved and bubble-dew point calculations were performed with association constants, using the method of Hayden and O'Connell.[22]

The optimization algorithm used for the data fitting was the Levenberg-Marquardt[23,24] technique, which is a local optimization method. It is, therefore, strongly dependent on the initial guesses for the parameters. In order to ensure that the best parameter values are obtained, the following scheme, involving sequential and simultaneous parameter estimation, using the results from the previous steps as initial guesses for each addition of a group/data set to the system, was employed. Also, to ensure that the parameters did not move too far away from the previous step, and, thus, guarantee that the previously investigated groups/data sets would still be represented with comparable accuracy, parameter regulariza-

Table 4. Methanol CI-Interaction Parameters

CI-interaction Parameter (i-j)	Value	CI-interaction Parameter (j-i)	Value
b_{C-C}	417.10	$\overline{b_{C-C}}$	14.01
c_{C-C}	−22.34	$\overline{c_{C-C}}$	−236.84
d_{C-C}	−54.19	$\overline{d_{C-C}}$	−250.59
e_{C-C}	−3.51	$\overline{e_{C-C}}$	0
b_{C-O}	396.91	$\overline{b_{C-O}}$	−94.79
c_{C-O}	−29.41	$\overline{c_{C-O}}$	−236.84
d_{C-O}	−94.63	$\overline{d_{C-O}}$	−250.59
e_{C-O}	−14.42	$\overline{e_{C-O}}$	0
b_{O-C}	−66.42	$\overline{b_{O-C}}$	86.40
c_{O-C}	61.12	$\overline{c_{O-C}}$	−97.14
d_{O-C}	−54.19	$\overline{d_{O-C}}$	−35.20
e_{O-C}	−3.51	$\overline{e_{O-C}}$	0
b_{O-O}	−74.84	$\overline{b_{O-O}}$	0.14
c_{O-O}	93.49	$\overline{c_{O-O}}$	−97.14
d_{O-O}	−94.63	$\overline{d_{O-O}}$	−35.20
e_{O-O}	−14.42	$\overline{e_{O-O}}$	0

Table 5. Water CI-Interaction Parameters

CI-interaction Parameter (i-j)	Value	CI-interaction Parameter (j-i)	Value
b_{C-C}	0	$\overline{b_{C-C}}$	0
c_{C-C}	0	$\overline{c_{C-C}}$	0
d_{C-C}	0	$\overline{d_{C-C}}$	0
e_{C-C}	0	$\overline{e_{C-C}}$	0
b_{C-O}	−845.17	$\overline{b_{C-O}}$	−2205.53
c_{C-O}	0	$\overline{c_{C-O}}$	0
d_{C-O}	0	$\overline{d_{C-O}}$	0
e_{C-O}	0	$\overline{e_{C-O}}$	0
b_{O-C}	−268.83	$\overline{b_{O-C}}$	123.53
c_{O-C}	706.28	$\overline{c_{O-C}}$	−1071.60
d_{O-C}	0	$\overline{d_{O-C}}$	0
e_{O-C}	16.30	$\overline{e_{O-C}}$	1.17
b_{O-O}	963.82	$\overline{b_{O-O}}$	909.49
c_{O-O}	71.54	$\overline{c_{O-O}}$	5821.92
d_{O-O}	0	$\overline{d_{O-O}}$	0
e_{O-O}	3.92	$\overline{e_{O-O}}$	0

tion was performed by adding a second term in the objective function.

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_{\text{exp}} - P_{\text{calc}}}{P_{\text{exp}}}\right)^2 + w_{reg}\sum_j \left(AIP_j - AIP_j^{\text{IG}}\right)^2 \quad (28)$$

where N is the number of experimental data points, AIP_j is the current value of the CI-interaction parameter j, AIP_j^{IG} its corresponding initial guess, and w_{reg} a weighting value used to increase or decrease the influence of regularization in the optimization.

Model Analysis and Prediction

Results and model analysis

The regressed CI-interaction parameters needed to predict a group-interaction parameter, are given for the "functional" groups in Table 3, for methanol in Table 4, and for water in Table 5.

Table 6 presents the average absolute relative deviation (AARD%; calculated using Eq. 29 for the reference UNIFAC method, and the CI-UNIFAC method over the entire data-set used for the estimation of the CI-interaction parameters presented in Tables 3–5, for the groups with C, H, O atoms and molecular groups, water and methanol. For correlation results involving also groups with N atoms, see Figure 4.

$$AARD(\%) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_{\text{exp}} - P_{\text{calc}}}{P_{\text{exp}}}\right| \times 100 \quad (29)$$

where N is the number of data points in the whole set of data.

From the results in Table 6, it can be noted that the average absolute relative deviation, AARD(%), is slightly lower for the CI-UNIFAC method when compared for the "functional groups"(2.54% vs. 2.57%), methanol (2.56% vs. 3.98%), and water (3.34% vs. 3.71%), even though less parameters (136 instead of 308 in the reference UNIFAC table[10]) are used to estimate VLE for the corresponding functional groups, methanol and water.

Figures 3–6 present the comparison between the deviation from experimental data with the original UNIFAC method, and the CI-UNIFAC method, for each of the data sets used in the parameter estimation process, and for each of the CI-interaction parameters sets (functional groups, methanol and water). The deviation plotted in these figures is the AARD (%) for the corresponding data set, defined by Eq. 29, but where N is the number of data points in the considered data set only. From these figures, one can see that the range of the deviation compared to the experimental data using the CI-interaction parameters to calculate group-interaction parameters is within the range of the deviation of the reference UNIFAC method. The deviations are even better in most cases, especially for the systems that are poorly represented with reference UNIFAC method. It indicates that the regressed CI-interaction parameters from Tables 3–5 can be used to calculate the missing group-interaction parameters, and will make VLE predictions with similar accuracy as the original UNIFAC method. Note that since the CI-UNIFAC method is proposed only as an extension of the reference UNIFAC method when the needed group interactions are missing, it is not reasonable to expect this method to correct the errors of the reference model. The KT-UNIFAC[25] has been developed to correct for some of the inaccuracies of the original UNIFAC method. However, based on the parameters (Tables 3–5) and the corresponding correlation results (Figures 3–6), it is reasonable to consider that the CI-UNIFAC method can be used for estimating missing group-interactions with acceptable accuracy. This certainly would be true if only one pair of missing UNIFAC group interactions need to be generated through the CI-UNIFAC method.

The interested reader can obtain a copy of the CI-interaction parameter tables, as well as the filled reference UNIFAC parameter table from the corresponding author. Also, more details on the numerical values related to Figures 3–6 can be obtained from the corresponding author.

Table 6. Average Absolute Relative Deviation (AARD) in % Over the Whole Set of Data

	C, H & O Atoms		C, H, O, & N Atoms	
	Reference UNIFAC[10]	CI-UNIFAC Method	Reference UNIFAC[10]	CI-UNIFAC Method
Functional groups	2.57	2.54	5.18	3.66
	Figure 3		Figure 4	
Methanol	3.98	2.56	–	–
	Figure 5			
Water	3.71	3.34	–	–
	Figure 6			

Comparison of the values for functional groups, methanol and water using the reference UNIFAC method,[10] and the CI-UNIFAC method.

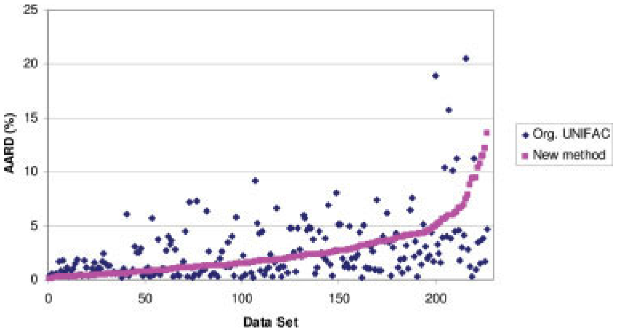

Figure 3. Functional groups: AARD (%) for each individual data set.

Comparison between the reference UNIFAC method,[10] and the CI-UNIFAC (new) method (results sorted by ascending deviation using the new method for groups with C, H & O atoms). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

VLE Prediction using the new method

The use of the CI-interaction parameters to calculate a missing group-interaction parameter requires the following:

(1) Use the water CI-interaction parameters (Table 5) to calculate any group-interaction parameter involving water,

(2) Use the methanol CI-interaction parameters (Table 4) to calculate any group-interaction parameter involving methanol, excluding methanol-water,

(3) Use the "functional groups" CI-interaction parameters (Table 3) to calculate any group-interaction, excluding the water and methanol interactions.

To estimate the group-interaction parameter between two groups k and l:

- If k has a lower group index than l, it will use the first set of CI-interaction parameters (b_{C-C}, b_{C-O},...) for the group-interaction parameter a_{kl}, and the second one ($\overline{b_{C-C}}$, $\overline{b_{C-O}}$,...) for the group-interaction parameter a_{kl},
- If k has a higher group index than l, it will use the second set of CI-interaction parameters ($\overline{b_{C-C}}$, $\overline{b_{C-O}}$,...) for the group-interaction parameter a_{kl}, and the first one (b_{C-C}, b_{C-O},...) for the group-interaction parameter a_{lk}.

The application of the CI-UNIFAC method is illustrated below through three case studies:

(1) Involves a system where an available group-interaction parameter is considered as missing, and calculated using the CI-interaction parameters,

(2) Involves a system where there is a missing group-interaction parameter, which is calculated using the CI-interaction parameters,

(3) Involves a system that is not well represented by the original UNIFAC groups, in which a new group is created, and its group-interaction parameters are calculated using the CI-interaction parameters.

For these examples, and whenever possible (cases (1) and (3)), a comparison of the deviation from experimental data is made between the original UNIFAC method, and the CI-UNIFAC method. Extensive tests have been made for the CI-UNIFAC method. However, when data is missing, it is difficult to provide any comparisons. Also, the three case studies follow the general trend of all the tests made for the CI-UNIFAC method.

Case study 1: Known Interaction Considered as Missing and Re-Estimated. In this case study, the CI-UNIFAC method is be used to re-estimate the interaction parameter between the CH_2 and the OH main-groups. Since these two groups have limited connectivity (only one nonhydrogen atom in each group), only the zeroth-order connectivity indices are used to calculate the group interactions. Also, since the CH_2 group has an index value of 1, and the OH group has an index value of 5, the b_{C-O} CI-interaction parameter will be used to calculate the a_{CH_2-OH} interaction, while the $\overline{b_{O-C}}$ atom interaction parameter will be used to calculate the a_{OH-CH_2} interaction. All the CI-interaction parameters used in the calculations are taken from Table 3. Details of the calculations are given below

$$a_{CH_2-OH} = b_{C-O}\frac{n_C^{(CH_2)}\chi_0^{(OH)} - n_O^{(OH)}\chi_0^{(CH_2)}}{\chi_0^{(OH)}\chi_0^{(CH_2)}} \tag{30}$$

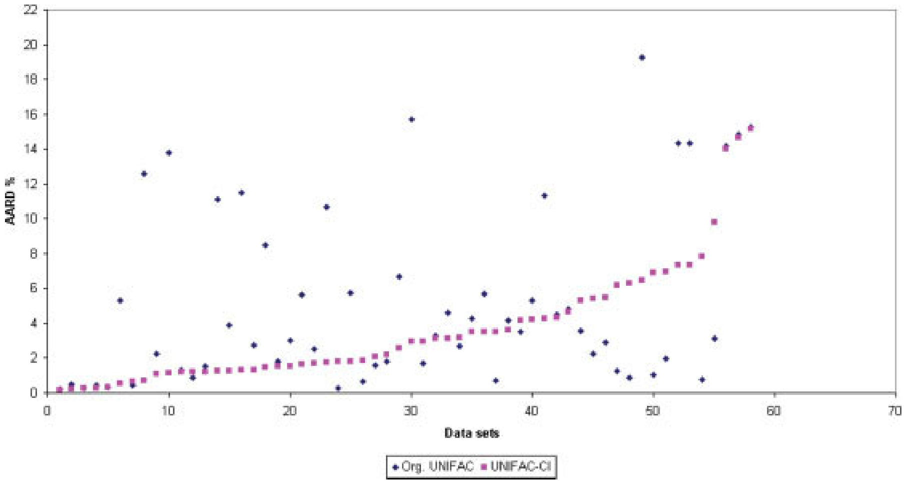

Figure 4. Functional groups: AARD (%) for each individual data set.

Comparison between the reference UNIFAC method,[10] and the CI-UNIFAC (new) method (Results sorted by ascending deviation using the new method for groups with C, H, O and N atoms for the 58 additional datasets). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

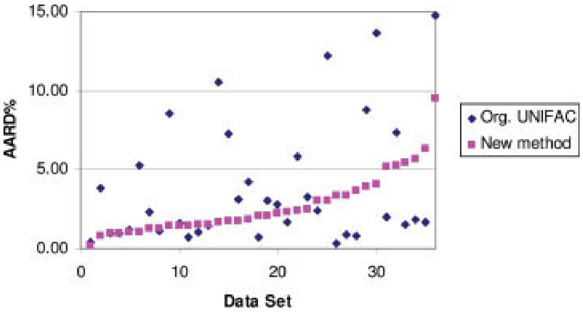

Figure 5. Methanol: AARD (%) for each individual data set.

Comparison between the reference UNIFAC method[10] and the CI-UNIFAC (new) method (Results sorted by ascending deviation using the new method for groups with C, H and O atoms). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$a_{OH-CH_2} = \overline{b_{O-C}}\,\frac{n_O^{(OH)}\chi_0^{(CH_2)} - n_C^{(CH_2)}\chi_0^{(OH)}}{\chi_0^{(CH_2)}\chi_0^{(OH)}} \qquad (31)$$

Applying Eqs. 30–31, and the connectivity index values for the groups from Table 2, we get, for a_{CH_2-OH} interaction

$$a_{CH_2-OH} = -1134.18 \times \frac{1 \times 0.44721 - 1 \times 0.70711}{0.44721 \times 0.70711}$$

$$a_{CH_2-OH} = 932.16$$

and for a_{OH-CH_2} interaction

$$a_{OH-CH_2} = 63.32 \times \frac{1 \times 0.70711 - 1 \times 0.44721}{0.70711 \times 0.44721}$$

$$a_{OH-CH_2} = 52.04$$

Table 7 presents the different values of the group-interaction parameters between the CH_2, and the OH groups with the reference UNIFAC method, and the CI-UNIFAC method.

The results of the two methods for predicting the vapor-liquid equilibrium are compared with an experimental data set[26] (2-butanol/n-hexane at 333.15 K), which was not used during the parameter estimation process. Results of this comparison are illustrated in Figure 7.

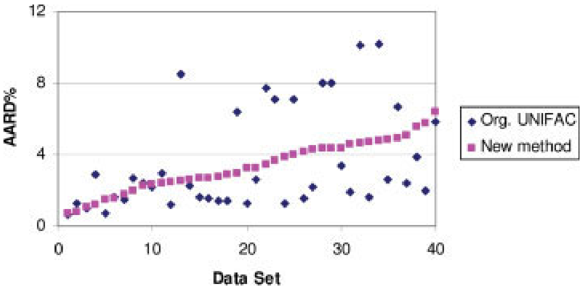

Figure 6. Water: AARD (%) for each individual data set.

Comparison between the reference UNIFAC method[10] and the CI-UNIFAC (new) method (Results sorted by ascending deviation using the new method for groups with C, H and O atoms). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 7. Group Interaction Matrix for CH2 and OH with the Reference UNIFAC Method[10] and the CI-UNIFAC Method

	Reference UNIFAC[10]		CI-UNIFAC Method	
Groups	CH_2	OH	CH_2	OH
CH_2	0	986.5	0	932.16
OH	156.4	0	52.04	0

As one can see from Figure 7, the results from the original UNIFAC method and the CI-UNIFAC method are very comparable for predicting the vapor-liquid equilibrium of the 2-butanol/n-hexane system. The new method is even slightly better near the azeotrope composition. Note, however, that a different set of UNIFAC group parameters[11] may give an improved UNIFAC correlation.

Case study 2: Estimation of a Missing Group-Interaction Parameter. In this case study, the CI-UNIFAC method is used to estimate the interaction parameter between the HCOO (formate), and the H_2O (water) groups, which is a missing interaction in the reference UNIFAC method[10] used in this article. Since the H_2O group has limited connectivity (only one nonhydrogen atom), only its zeroth-order connectivity index is used to calculate the group interactions, while the HCOO group has extended connectivity (three nonhydrogen atoms), and its zeroth-, first- and second-order connectivity indices will be used. Also, since the H_2O group has an index value of seven, and the HCOO group has an index value of 12, the b_{O-C}, c_{O-C}, e_{O-C}, b_{O-O}, c_{O-O} and e_{O-O} CI-interaction parameters will be used to calculate the $a_{H_2O-HCOO}$ interaction, while the $\overline{b_{C-O}}$ and $\overline{b_{O-O}}$ atom interaction parameters will be used to calculate the a_{HCOO-H_2O} interaction. Note that all the CI-interaction parameters used in the calculations need to be extracted from Table 5, since water is involved in the interactions. In this case, Eqs. 20–21 take the form (for purposes of illustration, only the first expression for the calculations and the final numerical results

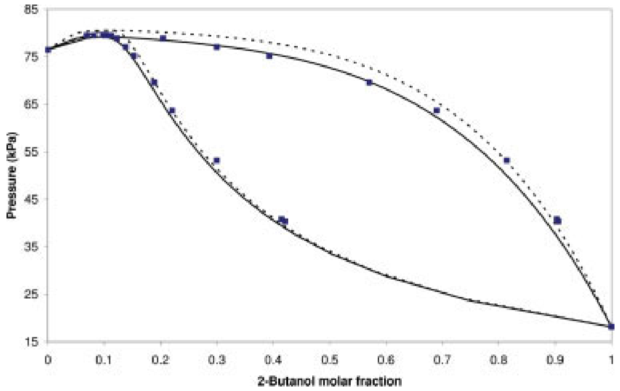

Figure 7. Comparison of the original UNIFAC method (dashed line), and the CI-UNIFAC method (full line) toward a 2-butanol/hexane experimental data set[26] (■) at 333.15 K.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 8. Group Interaction Matrix for CH_2, H_2O and HCOO with the Reference UNIFAC[10] Method and the CI-UNIFAC Method

	Reference UNIFAC[10]			CI-UNIFAC Method		
Groups	CH_2	H_2O	HCOO	CH_2	H_2O	HCOO
CH_2	0	1318	507	0	1318	507
H_2O	300	0	0	300	0	−256.75
HCOO	329.3	0	0	329.3	2314.75	0

are given)

$$\begin{aligned}a_{\mathrm{H_2O-HCOO}} &= b_{\mathrm{O-C}}\frac{n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}-n_{C}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_0^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ &+ C_{\mathrm{O-C}}\frac{n_{\mathrm{O}}^{\mathrm{H_2O}}\chi_1^{(\mathrm{HCOO})}-n_{\mathrm{C}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_1^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ &+ e_{\mathrm{O-C}}\frac{n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_2^{(\mathrm{HCOO})}-n_{\mathrm{C}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_2^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ &+ b_{\mathrm{O-O}}\frac{n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}-n_{\mathrm{O}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_0^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ &+ C_{\mathrm{O-O}}\frac{n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_1^{(\mathrm{HCOO})}-n_{\mathrm{O}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_1^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ &+ e_{\mathrm{O-O}}\frac{n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_2^{(\mathrm{HCOO})}-n_{\mathrm{O}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}{\chi_2^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}}\\ a_{\mathrm{H_2O-HCOO}} &= -256.75\end{aligned}$$

$$\begin{aligned}a_{\mathrm{HCOO-H_2O}} &= \overline{b_{\mathrm{C-O}}}\frac{n_{\mathrm{C}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}-n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}}{\chi_0^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}}\\ &+ \overline{b_{\mathrm{O-O}}}\frac{n_{\mathrm{O}}^{(\mathrm{HCOO})}\chi_0^{(\mathrm{H_2O})}-n_{\mathrm{O}}^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}}{\chi_0^{(\mathrm{H_2O})}\chi_0^{(\mathrm{HCOO})}}\\ a_{\mathrm{HCOO-H_2O}} &= 2314.75\end{aligned}$$

Table 8 lists the group-interaction parameter table involving the CH_2, H_2O and HCOO main-groups with the original UNIFAC method (where the interactions between H_2O and HCOO are missing), and the CI-UNIFAC method (where the interactions between H_2O and HCOO are calculated).

The results of the calculations and the use of the CI-UNIFAC method for predicting the vapor-liquid equilibrium is assessed through a set of experimental data[27] (ethyl formate/water at 300 K), which was not used during the CI-interaction parameter estimation process. Results of this comparison are illustrated in Figure 8.

As one can see from Figure 8, the CI-UNIFAC method, by allowing the user to calculate the interaction between H_2O and HCOO, permits to estimate the vapor-liquid equilibrium of the ethyl formate/water with a reasonable quality. Note that the azeotropic point is predicted quite well, even though the deviations in pressure from the experimental values are quite high for ethyl formate mole fractions less than 0.1. This, however, may be an error from the original UNIFAC group-interaction parameters, which the CI-interactions are not attempting to correct. That is, using additional data, the water-CH_2, water-HCOO and CH_2-HCOO interactions could be fine tuned or re-estimated together with all the other group interactions. The missing UNIFAC group parameters are available in Wittig et al.[11] which also gives an improved prediction of the VLE.

Case Study 3: Creation of Group and Estimation of its Group-Interaction Parameters. In case study 3, the new method is used to create a group and its group-interaction parameters. In the original UNIFAC method, 1,2-ethanediol is described as the combination of two CH_2 groups and two OH groups. In this example, 1,2-ethanediol (ethylene glycol) will be considered as a group by itself (noted ETHA). We will use the new method for estimating the vapor-liquid equilibrium between 1,2-ethanediol (now considered as a group by itself) and methanol.

For calculating the corresponding group-interaction parameters, one needs first to calculate its connectivity indices. According to the method described before, its zeroth-, first- and second-order connectivity indices are respectively 2.31, 1.13 and 0.45. Also, according to the rules of Bondi,[17] the corresponding surface Q_k and volume R_k parameters are equal to 3.02 and 2.92, respectively.

Since methanol only has zeroth- and first-order connectivity indices, the $b_{\mathrm{C-C}}$, $c_{\mathrm{C-C}}$, $d_{\mathrm{C-C}}$, $e_{\mathrm{C-C}}$, $b_{\mathrm{C-O}}$, $c_{\mathrm{C-O}}$, $d_{\mathrm{C-O}}$, $e_{\mathrm{C-O}}$, $b_{\mathrm{O-C}}$, $c_{\mathrm{O-C}}$, $d_{\mathrm{O-C}}$, $e_{\mathrm{O-C}}$, $b_{\mathrm{O-O}}$, $c_{\mathrm{O-O}}$, $d_{\mathrm{O-O}}$ and $e_{\mathrm{O-O}}$ CI-interaction parameters will be used to calculate the $a_{\mathrm{CH_3OH-ETHA}}$ interaction, while the $\overline{b_{\mathrm{C-C}}}$, $\overline{c_{\mathrm{C-C}}}$, $\overline{d_{\mathrm{C-C}}}$, $\overline{b_{\mathrm{C-O}}}$, $\overline{c_{\mathrm{C-O}}}$, $\overline{d_{\mathrm{C-O}}}$, $\overline{b_{\mathrm{O-C}}}$, $\overline{c_{\mathrm{O-C}}}$, $\overline{d_{\mathrm{O-C}}}$, $\overline{b_{\mathrm{O-O}}}$, $\overline{c_{\mathrm{O-O}}}$ and $\overline{d_{\mathrm{O-O}}}$ CI-interaction parameters will be used to calculate the $a_{\mathrm{ETHA-CH_3OH}}$ interaction. Note that all the CI-interaction parameters used in the calculations need to be extracted from Table 4, since methanol is involved in the interactions. Using Eqs 21 to 25 like in other case studies, the $a_{\mathrm{CH_3OH-ETHA}}$ and $a_{\mathrm{ETHA-CH_3OH}}$ are calculated and found to be −232.8 and 1184.6, respectively.

The values of the group-interaction parameters (needed for the binary system) for the original UNIFAC method are com-

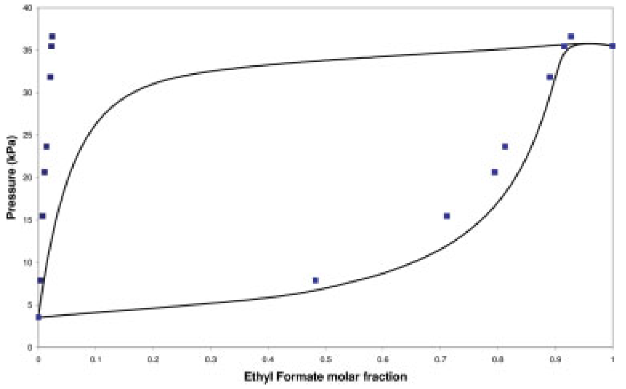

Figure 8. Assessment of the results given by the CI-UNIFAC method (full line) for calculating a missing UNIFAC group-interaction parameter towards a ethyl fromate/water experimental data set[27] (■) at 300 K.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 9. Group Interaction Matrix for CH_2, OH and CH_3OH using the Reference UNIFAC[10] Method, and CH_3OH and the new ETHA Group using the CI-UNIFAC Method

	Reference UNIFAC[10]				CI-UNIFAC Method	
	CH_2	OH	CH_3OH		CH_3OH	ETHA
CH_2	0	986.5	697.2	CH_3OH	0	−232.8
OH	156.4	0	−137.1	ETHA	1184.6	0
CH_3OH	16.51	249.1	0			

pared with those predicted by the CI-UNIFAC method in Table 9.

The performance of the CI-UNIFAC method for the prediction of VLE involving the system 1,2-ethanediol/methanol at 760 mm Hg, has been compared with a set of experimental data,[28] which was not used during the parameter estimation process. Results of this comparison are illustrated in Figure 9.

As one can see from Figure 9, the predicted VLE with the original UNIFAC method (with 1,2-ethanediol described as a combination of CH_2 and OH groups), and the new method (with 1,2-ethanediol being a group by itself) are very comparable for predicting the VLE for this binary system. The CI-UNIFAC method is even slightly better for 1,2-ethanediol mole fractions between 0 and 0.5.

In Figure 10a, b, the performance of the CI-UNIFAC method is highlighted where all the group interactions have been estimated through the CI-interaction parameters. Figure 10a highlights the VLE for the binary mixture of Butyronitrile/2-butanol at 298.15 K,[29] while figure 10b highlights the VLE for the binary mixture of aniline/ethyl-acetate at 348.15 K.[30]

Conclusions

The CI-UNIFAC method for estimating missing UNIFAC group interaction parameters provides a simple means for calculation of vapor-liquid equilibrium compositions in the frequently encountered situation, where some UNIFAC group-interaction parameters are not available. CI-interaction parameters that can be used to calculate missing pairs of group-interaction parameters are given for the case of UNIFAC functional (main) groups with C, H, O and N atoms,

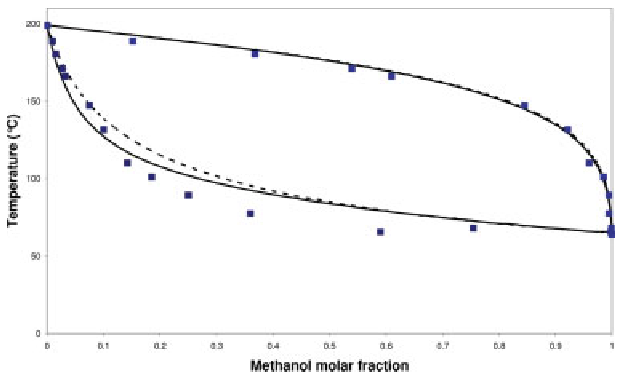

Figure 9. Comparison of the original UNIFAC method (dashed line), and the CI-UNIFAC method (full line) toward a 1,2-ethanediol/methanol experimental data set[28] (■) at 760 mm Hg.

[Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

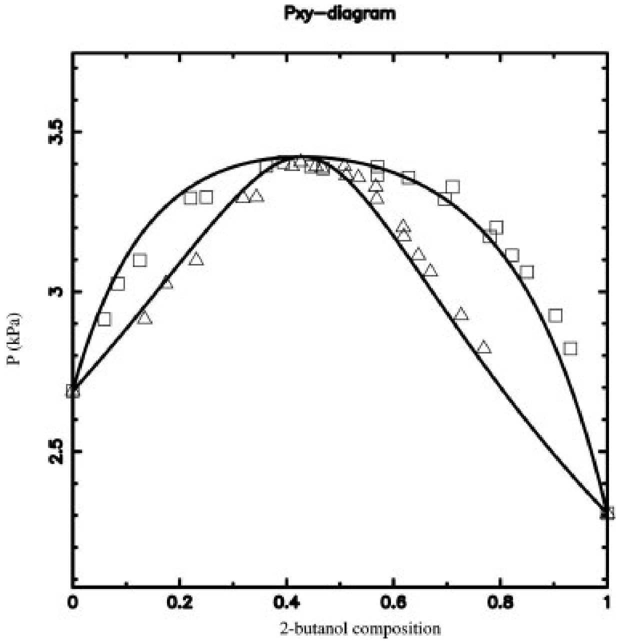

Figure 10a. VLE calculations with the CI-method vs. experimental data[29] for butyronitrrile/2-butanol at 298.15 K (Δ).

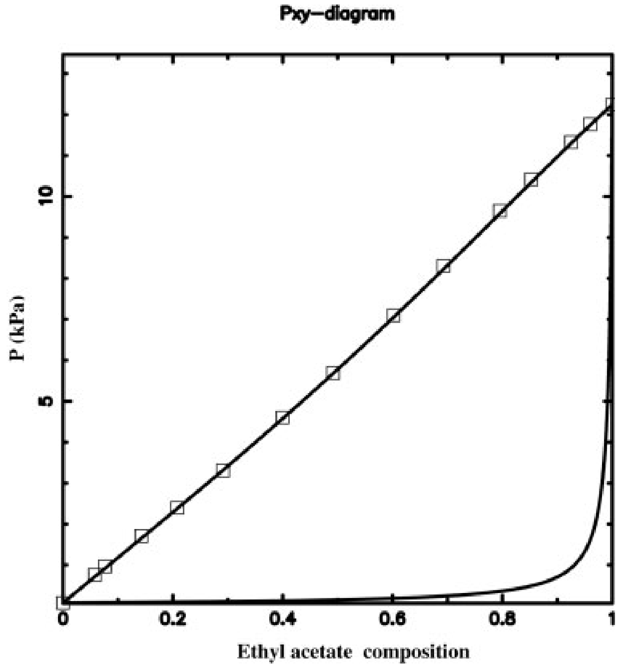

Figure 10b. VLE calculations with the CI-method vs. experimental data[30] for aniline/ethyl-acetate at 298.15 K (Δ).

and, molecular-groups methanol and water. The accuracy of prediction with the CI-UNIFAC method has been found to be nearly the same as that of the reference UNIFAC method. The performance of the CI-UNIFAC method has been highlighted through three representative case studies involving the calculation of group-interaction parameters and prediction of liquid phase activity coefficients, when known group-interactions are considered as missing, there is a missing pair of group-interactions for the original UNIFAC method, and when a new group is created and its interaction predicted. In all cases, the predicted vapor-liquid equilibrium has been found to be acceptable and compared well with experimental data and predictions obtained using the original UNIFAC method (when possible). While the present range of applicability is limited to components containing only C, H, O and N atoms, and to vapor-liquid equilibrium, this range can easily be expanded to additional atoms and to different types of phase equilibrium. We believe the scope of the CI-UNIFAC method is very large as it allows the creation of a very large number of groups and their parameters. To keep the CI-UNIFAC method manageable, it is recommended not to define UNIFAC groups with more than four atoms, out of which two should be C and H atoms. Current and future work is extending the CI-interactions for groups involving S, followed by halides. At the same time, a parallel work is estimating the CI-interactions for LLE and evaluating the use of VLE/LLE parameters for SLE.

Notation

a_{kl} = UNIFAC group interaction parameter between group k and group l
$b_{X-Y}, \overline{b_{X-Y}}$ = 0th-order CI-interaction parameter between atom X and atom Y
$c_{X-Y}, \overline{c_{X-Y}}$ = 1st-order CI-interaction parameter between atom X and atom Y
$d_{X-Y}, \overline{d_{X-Y}}$ = 2nd-order CI-interaction parameter between atom X and atom Y
$e_{X-Y}, \overline{e_{X-Y}}$ = 2rd-order CI-interaction parameter between atom X and atom Y
G_{ki} = defined in Eq. 15
J_i = defined in Eq. 11
L_i = defined in Eq. 12
N_H = number of hydrogen attached to the atom
$n_X^{(k)}$ = number of atoms of type X in group k
P = pressure
P^{sat} = vapor pressure
Q_k = surface parameter, for group k
q_i = molecular surface parameter, for component i
r_i = molecular volume parameter, for component i
R_k = volume parameter, for group k
S = value of the objective function
s_{ki} = defined in Eq. 17
T = temperature
x_i = mole fraction of component i in the liquid phase
w_{reg} = weighting factor for regularization
Z_v = number of valence electron in the atom
$\check{\mathrm{Z}}$ = atomic number

Greek letters

δ^v = 0th-order (atomic) valence index
β^v = 1st-order (bond; path of length one edge) valence index
ε^v = 2nd-order (path of length two edges) valence index
γ_i = activity coefficient for component i
γ_i^{COM} = activity coefficient, combinatorial part
γ_i^{RES} = activity coefficient, residual part
θ_k = defined in Eq. 16
η_k = defined in Eq. 18
$v_k^{(i)}$ = number of groups k present in molecule i
τ_{mk} = defined in Eq.19
${}^v\chi_m^{(k)}$ = valence connectivity index of order m in group (or molecule) k

Indexes

calc = calculated quantity
COM = combinatorial part
exp = experimental quantity
IG = initial guess
RES = residual part

Literature Cited

1. Gani R, Harper P, Hostrup M. Automatic creation of missing groups through connectivity index for pure component property prediction. *Ind & Eng Chem Res.* 2005;44:8436–8454.
2. Fredenslund A, Gmehling J, Rasmussen P. Vapor-Liquid Equilibria Using UNIFAC; Amsterdam: Elsevier; 1977.
3. Magnussen T, Rasmussen P, Fredenslund A. UNIFAC parameter table for prediction of liquid-liquid equilibria. *Ind & Eng Chem Fund.* 1981;20:331–339.
4. Gmehling JG, Anderson TF, Prausnitz JM. Solid-Liquid Equilibria Using UNIFAC. *Ind & Eng Chem Fundamental.* 1978;17:269–273.
5. Oishi T, Prausnitz JM. Estimation of solvent activities in polymers using a group contribution method. *Ind & Eng Chem Proc Des & Develop* 1978;17:335–339.
6. Jensen T, Fredenslund A, Rasmussen P. Pure Component Vapor Pressures Using UNIFAC group Contribution. *Ind & Eng Chem Fund.* 1981;20:239–246.
7. Gmehling J, Rasmussen P. Flash points of flammable liquids using UNIFAC. *Ind & Eng Chem Fund.* 1982;21:186–188.
8. Sander B, Skjold-Jørgensen S, Rasmussen P. Gas solubility calculations. I. UNIFAC. *Fluid Phase Equilibria.* 1983;11:105–126.
9. Dang D, Tasslos DP. Prediction of enthalpies of mixing with a UNIFAC model. *Ind & Eng Chem Proc Des & Develop.* 1986;25:22–31.
10. Hansen HK, 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.
11. 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.
12. Kier LB, Hall HL. *Molecular Connectivity in Structure Activity Analysis.* New York: John Wiley & Sons, Inc.; 1986.
13. Kier LB, Hall HL. Issues in representation of molecular structure the development of molecular connectivity. *J of Mol Graphics & Modelling.* 2001;20:4–18.
14. Trinajstic N. *Chemical Graph Theory.* Ohio: CRC Press; 1983.
15. Randic M. *Connectivity Index 25 Years After. J of Molecular Graphics & Modelling.* 2001;20:19–35.
16. Smith JM, Van Ness HC. *Introduction to Chemical Engineering Thermodynamics*, 4th ed. New York: Mc Graw Hill; 1996.
17. Bondi A. *Physical Properties of Molecular Crystals, Liquids and Glasses*; New York: John Wiley & Sons; 1968.
18. Van Ness HC, Byer SM, Gibbs RE, Part I. An appraisal of data reduction methods. *AIChE J.* 1973;19:238–244.
19. Nielsen TL, Abildskov J, Harper PM, Papaeconomou I, Gani R. The CAPEC Database. *J of Chem Eng Data.* 2001;46:1041–1044.
20. e-thermo® Online Database, Processium, Villeurbanne, France. www.processium.com
21. Van Ness HC. Thermodynamics in the treatment of (vapor + liquid) equilibria. *J Chem Thermodynamics.* 1995;27:113–134.
22. Hayden JG, O'Connell JP. A generalized method for predicting second virial coefficients. *Ind & Eng Chem Proc Des & Develop.* 1975;14:209–216.
23. Levenberg KA. Method for the solution of certain problems in least squares. *Quarterly App Math.* 1944;2:164–168.
24. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J of App Math.* 1963;11:431–441.

25. Kang JW, Gani R, Abildskov J, Cobas J. Estimation of mixture properties from first and second-order group contributions with UNIFAC models. *Ind & Eng Res*. 2002;41:3260–3273.
26. Maczinska Z. *International data series. Selected Data Mixtures*. 1983;3:199–201.
27. Linek J, Wichterle I, Marsh KN. Vapor-liquid equilibria for water + diacetone alcohol, ethyl methanoate + water, and ethyl methanoate + phenol. *J Chem Eng Data*. 1996;41:1219–1222.
28. Baker TH, Fisher GT, Roth JA. Vapor liquid equilibrium and refractive indices of the methanol-ethylene glycol system. *J Chem Eng Data*. 1964;1:11–12.
29. Garriga R, Sanchez F, Perez P, Gracia M. Excess gibbs energies at eight temperatures and excess enthalpies and volumes at 298.15 K for butanenitrile plus 2-Butanol. *J Chem Eng Data*. 1997;42:78–83.
30. Khurma JR, Muthu O, Munjai S, Smith BD. Total–pressure vapor-liquid equilibrium data for binary systems of aniline with 1-chlorobutane and ethyl-acetate. *J Chem Eng Data*. 1983;28:108–112.

Manuscript received Jun. 21, 2006, revision received Dec. 22, 2006, and final revision received Mar. 2, 2007.