THE DEVELOPMENT OF A HYBRID ACTIVITY COEFFICIENT MODEL UTILIZING THE SOLUTION OF GROUPS CONCEPT

By

Brian J. Satola

[BSc (Eng.) / Chemical Engineering; University of Akron – Ohio, U.S.A.]

University of KwaZulu-Natal, Howard College Campus,

For the degree of Master's of Science (Chemical Engineering)

March 2011

ABSTRACT

During the course of this thesis the UNIFAC method (group-based method) was regressed to individual Px(T) binary datasets, and the results are compared to the regression results using the Wilson, NRTL, and UNIQUAC equations (component-based models). It is shown that these component-based methods best represent the experimental data when the comparisons are restricted to those systems defined by only two UNIFAC maingroups. For those systems requiring three or more maingroups, however, the regressions using the UNIFAC method (i.e. the group-based approach) are shown to provide the best reproducible results.

Evaluations are also presented on the ability of the UNIFAC and mod. UNIFAC (Do.) methods to reproduce experimental activity coefficients at infinite dilution for single and co-solvent systems. For the case of single solvent-systems the newly developed MRR combinatorial expression (Moller, 2010) is evaluated as a direct combinatorial replacement for both methods, although it was originally developed only for estimating activity coefficients at infinite dilution in alkane-solvents. Overall, it is shown that the best results are obtained using the mod. UNIFAC (Do.) method, and that poor results are obtained when trying to use the MRR combinatorial as a direct combinatorial replacement in either method (for systems other than alkane-solvents).

Given the favourable results obtained using the mod. UNIFAC (Do.) method, the model was used to generate pseudo data points at multiple temperatures for regression using the NRTL equation, where parameters quadratic in temperature were fitted. It is shown that one may introduce unnecessary errors when translating these predictions into the model parameters of the NRTL equation. In order to eliminate these potential "losses in translation," a new liquid activity coefficient model/methodology is being proposed.

Instead of using group contribution methods as second-choice data generators, it is proposed that these predictive methods be employed in a more direct fashion in process simulations. Instead of regressing experimental data using component-based methods such as NRTL and Wilson, the error in the predicted results are regressed by layering one of these methods on top of a group contribution method like mod. UNIFAC (Do.). This is the fundamental idea behind the proposed *hybrid methodology/models*.

Results are presented for two hybrid models, where the NRTL and Wilson equations are used to correct for the predictions made using the mod. UNIFAC (Do.) method. These methods are being called NRTL-FAC(Do.) and Wilson-FAC(Do.) respectively. In most cases, it is shown that the overall regression results using these new models are as good as or better than the individual models making them up.

All experimental data used in this dissertation was obtained from the Dortmund Data Bank (DDBST Software and Separation Technology GmbH, 2009), and all predictions made using the UNIFAC and mod. UNIFAC (Do.) methods were calculated using the Consortium parameters (The UNIFAC Consortium, 2008).

PREFACE

The work presented in this thesis report was performed at the University of KwaZulu-Natal, Durban from February 2009 to December 2010. The work was supervised by Prof. D. Ramjugernath and Prof. Dr. J. Rarey.

As the candidate's supervisor, I, Prof. D. Ramjugernath, have approved this thesis for submission.

____

Prof. D. Ramjugernath

Declaration

I ________________________ declare that:

(i.) The research reported in this thesis, except where otherwise indicated, is my original work.

(ii.) This thesis has not been submitted for any degree or examination at any other university.

(iii.) This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

(iv.) This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

- a.) their words have been re-written but the general information attributed to them has been referenced.
- b.) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.

(v.) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.

(vi.) This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the dissertation/thesis and in the References Chapters.

____

Brian J. Satola (209525595)

ACKNOWLEDGEMENTS

I would like to acknowledge the following people for their contribution to this work:

> My supervisor, Prof. Dr J. Rarey [Carl von Ossietzky University, Oldenburg, Germany] for his wealth of ideas, wisdom, support, and inspiration. Your good nature, easy smile, and zest for life have truly made a deep impact on my life.
>
> My supervisor, Prof. D. Ramjugernath for taking the chance on inviting me into his research group, and for his patience, guidance, and understanding throughout my study. Not to mention the additional developmental opportunities beyond what any student could expect to obtain while studying. I very much look forward to continuing my studies in your group.
>
> Thanks to all the friends that I have made here because of pursuing this degree in South Africa. Although I have had some scrapes here and there, I have found it to be a warm and inviting place. I have truly enjoyed the experience, and I look forward to enjoying it in the years to come.
>
> To my close friend Joseph McHugh, remember it is not too late to get out while you can. Once you say those magic words, it is a life sentence with no possibility for parole (well...). Joking aside, I wish you and Courtney all my best, for the future, and all it holds.
>
> My parents Chester and Linda and even my sister Krista, for supporting me in everything I choose to do in my life, even though these choices take me far from home more often than not. Thanks for putting up with me; I know I can be the source of much headache and trouble at times. I would not be where I am today if it was not for your nurturing, care, and support – I love you all more than words can say.

Table of Contents

List of Figures

List of Tables

Nomenclature

Latin Letters

$a, b, c, \cdots, A, B, C, \cdots$	parameters of equations
a	molar Helmholtz energy (J/mol)
A	Helmholtz energy (J)
$\mathbb{C}_i, \cdots Comp$	component *i*, in a solution of components
E	Energy (J)
f	fugacity
$f(x)$	a function of (x) variables
$Final$	final state
g	molar Gibbs energy (J/mol)
G	Gibbs energy (J)
h	molar Enthalpy (J/mol)
H	Enthalpy (J)
$\mathbb{G}_{\mathrm{i}}, \cdots Group$	molecular group *i,* in a solution molecular groups
$Initial$	initial state
k	Boltzmann constant
l_i, θ'_i, and Φ'_i	convenience variables used in UNIQUAC equation
LLE	liquid-liquid equilibria
M	any intensive thermodynamic property
nM	any extensive thermodynamic property
n, N	number of moles, data points, particles, or molecules
p_i^*	probability of quantum state i of a system's ensemble, where $\sum_i p_i^* = 1$ (summation over all possible quantum states)
P	pressure (kPa)
q	pure-component area parameter
Q	heat (J), or canonical partion function
Q_k	area parameter group/segment k
r	pure-component volume parameter
R	gas constant
R_k	volume parameter group/segment k
s	molar entropy (J/mol)

S	entropy (J)
SLE	solid-liquid equilibrium
T	absolute temperature
u	molar internal energy (J/mol)
U	internal energy (J)
v	molar volume (cm^3/mol)
V	volume (cm^3)
VLE	vapour-liquid equilibrium
w	interchange energy
W	work (J)
x	mole fraction (liquid)
X	group fraction in liquid solution
z	coordination number (number of nearest touching neighbours for a central molecule)

Greek Letters

α	nonrandomness parameter in NRTL equation
γ	activity coefficient
Γ	group activity coefficient
ε	interaction/potential energy
ξ	local volume fractions
ζ_{mk} and ξ_{mk}	group pair parameters characteristic of groups m and k, independent of temperature, used in ASOG method
θ	component surface area fraction
Θ	group surface area fraction
Φ_i'	component volume correction used in the combinatorial term of the modified UNIFAC method
Ψ_{mk} and T_{mk}	group interaction parameters between group m and group k (cal/mol)
Λ	binary interaction parameters in Wilson equation (cal/mol)
μ	chemical potential
τ	energetic parameters of the NRTL and UNIQUAC equations $(cal \cdot mol^{-1})$

Φ	volume fraction or segment fraction
ν_k	number of k molecular segments

Symbols and Numbers

$1, 2, 3, \cdots \# \cdots *$	system component or group numbers
C_P	heat capacity at constant pressure (J/kg)
$g(N_1, N_2, N_{12})$	represents the number of ways molecules of component 1 and 2 can be arranged on a lattice
k	Planck's constant $(6.62607 \times 10^{-34} m^2 kg \cdot s^{-1})$
N_A	Avogadro's number $(6.02214 \times 10^{23} mol^{-1})$
R	universal/ideal gas constant $8.314472\, J \cdot K^{-1} mol^{-1}$
$\mathrm{R_i^*}$	pure-component group ratio

Sub- and Superscripts: Latin

AZD	azeotropic value/point
C, or COM	combinatorial quantity
$Comp$	total number of system components
E	excess property
FH	Flory-Huggins
FV	free volume
i, j, k	components, or groups i, j, k
(i), (j)	hypothetical fluids, or pure components i and j
is	ideal solution property
L	liquid phase
(M), or (Mix)	mixture quantity (e.g. solution of groups)
p	phase p
$Phase$	total number of phases in system
rev	reversible
R, or RES	residual quantity

Sub- and Superscripts: Greek

α, β, π	phases α, β, and π

Sub- and Superscripts: Symbols and Numbers

0	standard state, or reference state
$1, 2, 3 \cdots$	system component identifiers
∞	infinite dilution
$\wedge$	component property
$-$	partial molar mixture property

1. Introduction

Models for the description of real non-electrolyte mixture behaviour are of fundamental importance for the synthesis, simulation, design, and operation of many separation processes used in industry (distillation and extraction are such examples). Since 60-80% of the total costs often arise in the separation step, a reliable knowledge of the phase equilibrium behaviour of the system to be separated is of special importance to industrial practitioners (Gmehling, 2009).

Models like Wilson, NRTL, and UNIQUAC have long been used with great success for the description of the real behaviour of multicomponent mixtures, but they are restricted due to the limited availability of binary interaction parameters. Out of this reason, group contribution methods like ASOG, UNIFAC, and mod. UNIFAC were developed that are based on structural group interactions instead of molecules. These predictive models however rely on *large* experimental datasets for regression purposes; once the group interaction parameters are determined, there is currently no way of adjusting the predicted results for a single component pair without potentially influencing the description of other binary systems.

1.1. Problem Statement

In several cases, it has been observed that group interaction models can represent binary mixtures more accurately if the parameters are regressed to experimental data on a case by case basis (Gmehling, 2009). For this reason one of the objectives was to test to what extent the group interaction concept is advantageous. The subsequent step and main focus of the project was then to investigate the potential benefit of a hybrid method, where results from group contribution methods could be adjusted by superimposing component-component interactions regressed to actual mixture data.

1.2. Solution Overview

General practice is to use predicted data (i.e. artificial) to fit missing model parameters where no experimental information is available. This allows the retention of the model parameters fitted to actual experimental data, while incorporating predictive results from group models like UNIFAC (i.e. GC2gE conversion, where results from a group contribution method are translated/regressed using a g^E-model). This means that predictive models are mainly being used indirectly in order to retain parameters that were fitted directly to experimental data using component models such as Wilson or NRTL (typically more accurate). This fundamental need to stay within the component-based approach can lead to a “loss in translation,” whereby the model fitted to predictive results performs worse than if the predictive model was used directly.

By introducing a tuneable portion to predictive models practitioners would be able to use the group method throughout their simulation. This would eliminate the need to fit component interaction models to predictive results (which could lead to a loss in translation), and would still allow the regression of binary parameters between individual components to any existing experimental data—an essential requirement for industrial applications.

1.3.Thesis Overview

A review of thermodynamic fundamentals was first performed to ensure that a strong foundation existed on which to build on. This then facilitated a literature review tracing the development of methods, models, and procedures used for the calculation of the real phase behaviour of binary and higher mixtures.

Special emphasis was given to models typically used in industry for low to moderate pressures, particularly those models based on the local composition concept (LC, or component-based models) and those based on the solution of groups concept (SOG, or group contribution models). These included the component-based models Wilson, NRTL, and UNIQUAC and the group contribution methods ASOG, UNIFAC, and mod. UNIFAC (Do.). Additional details concerning these methods are contained in the literature review of Chapter 3.

Available literature was also searched to verify the originality of the proposed hybrid method, which is first presented in Chapter 4, and then explained in further detail later in Chapter 5.5. To the best of the author's knowledge no such method, as described in this work, can be found in available public literature. It should be noted, however, that the idea of combining two different model-types is not new. It is often beneficial to combine models to achieve results which are impossible, difficult, and/or superior to those results obtained using each of the combined models separately.

Similar ideas, in fact, are employed every day in performing practical phase equilibrium calculations; the "Gamma-Phi" methodology is one such case. Where an equation of state (EOS) is used for defining the fugacity coefficient for the description of the vapour phase, and where an activity coefficient model is used to define activity coefficients used for the description of the liquid phase. For low to moderate pressures, this arrangement typically results in the best description of the real phase behaviour of mixtures. The non-idealities in the liquid phase are handled by an activity coefficient model (where many mixing rules and equations of state typically have trouble) and the non-idealities in the vapour phase are handled by an equation of state (which activity coefficient models do not describe/account for).

During the course of this work, it also became necessary to gain a familiarity with the calculated behaviour of mixtures using the component-based models and group contribution models typically used in industry. For comparison, this required calculations to be performed for hundreds of thousands of data points. In the case of component-based models such as Wilson, NRTL, and UNIQUAC this first required the fitting of model parameters to available mixture data, and then using these same parameters to re-evaluate the data they were fitted to. As for the group contribution methods UNIFAC[1] and mod. UNIFAC (Do.)[2], both the available public parameters and Consortium values (2008) were initially evaluated; however since a larger number of datasets could be calculated using the Consortium parameters and since the Consortium versions for these

---

[1] For public UNIFAC parameters: (Fredenslund, et al., 1975), (Fredenslund, et al., 1977), and (Hansen, et al., 1991).

[2] For public mod. UNIFAC (Do.) parameters: (Weidlich, et al., 1987), (Gmehling, et al., 1993), and (Jakob, et al., 2006).

methods are far more advanced, all results presented in this thesis/dissertation have been performed using these parameter values.

As implied thus far, large amounts of experimental data were required for testing purposes. These experimental values were obtained from the Dortmund Data Bank (DDB, (2009)), which can be considered one of the largest databases of thermophysical properties known to exist. The DDB software package (DDBSP, (2009)) was also used in conjunction with the DDB, which aided in the handling, manipulation, correlation, prediction, and calculation of various thermophysical properties stored within the databank, not to mention additional tools for process synthesis (e.g. GC2gE conversion). In order to use these programs effectively, however, a solid familiarity had to be gained. This was obtained through personal instruction (Rarey, 2009-2010), online training material (DDBST Software and Separation Technology GmbH, 2010), user-interactions, and experience gained through actual use of these programs.

Microsoft Office Suite 2007 (Microsoft, 2007) was also used extensively during the course of the investigation. Microsoft Access (Microsoft, 2007) was used to store the hundreds of thousands of experimental data points, and the numerous calculation results obtained from routines developed during the course of the project. Where Microsoft's Visual Basic for Applications (VBA) programming language (included in Microsoft Office Suite 2007) was used for the development of these calculation routines, using Microsoft Excel (Microsoft, 2007) as the graphical user interface (GUI). A large part of the work involved developing ways in which such large amounts of data could be handled, manipulated, and visualized.

The PEQ Thermo Engine of DDBSP (DDBST Software and Separation Technology GmbH, 2009) was used to make property calculations wherever possible, but it also became necessary to program a number of the thermophysical property models directly into VBA. After many trials and tribulations, an object-oriented approach was learned and adopted for the routines developed for this project. This approach allows for data to be used in a consistent manner. For example, a class-file (an object) was created to retrieve and handle pure component property parameters (named PCPAR). This means that whenever pure component property parameters are required in a routine, a simple reference to the PCPAR class/object is all that is needed, instead of having to handle the property files for each of the 25,000+ components separately each time a property-value is requred. The PCPAR class/object, in this case, can therefore be thought of as a box containing all of the parameters for the database of pure components that was created. The object approach therefore facilitates the reusability of common functions, methods, and properties inherent to these class-objects. Decreasing the development time for new routines, for example, which require information handled by the PCPAR class/object.

The development of the tools required to perform the necessary calculations, and statistical evaluations of the results was a considerable undertaking. For those with little to no programming experience, the amount of time and effort it takes to develop these software tools is hard to convey. Knowledge of various object models was needed in this case: the access database object module (ADO) for the storage and retrieval of project values, the PEQCOM library of DDB for flash calculations, and Office Web Components (OWC) for the charting of results, just to name a few. These objects were then used to create programs to handle, generate, manipulate, and to visualize the millions of data points resulting from the study.

All of these tools were instrumental in evaluating the component-based models, group contribution methods, and hybrid models covered in this report, where the ability of the group contribution approach to represent the real behaviour of mixtures is discussed in Chapter 5.1 (Evaluation of the Solution of Groups Concept), and the resulting details of which are discussed in Chapter 5.1.2. A new combinatorial expression developed (MRR combinatorial) has shown promising results for alkane solvent mixtures (Moller, 2010), and it was therefore desired to evaluate its potential use in the proposed hybrid method. This required the evaluation of the MRR as a direct replacement for the combinatorial expressions used by UNIFAC and mod. UNIFAC (Do.), although its development was intended for use in alkane solvent mixtures only. Since dilute mixtures often represent extreme changes in property values, the ability of these group contribution methods to predict activity coefficients at infinite dilution for binary and ternary mixtures were also evaluated (Chapter 5.2 and Chapter 5.3 respectively), including the variations of these methods using the MRR combinatorial as a direct replacement (systems of two components only).

After evaluating the pros and cons of these models, the “loss in translation” effect was evaluated. Since it was observed that the best overall performance could be obtained using the mod. UNIFAC (Do.) method, it was used for generating the pseudo data required for regression purposes, where the component-based model of NRTL (fitting parameters quadratic in temperature) was used. This process is referred to as performing a GC2gE conversion (DDBST Software and Separation Technology GmbH, 2010), which can lead to situations where the model fitted to the predictive results performs worse than if the predictive model was used directly. These results are presented in Chapter 5.4.

One way to completely remove this effect would be to use group predictions directly in process simulations, while providing practitioners a way to influence the calculated behaviour of specific component binaries. One way of doing this is to layer a component-based correction on top of a predictive model—the proposed hybrid activity coefficient model. This hybrid model is first introduced in Chapter 4, and is further discussed and evaluated in Chapter 5.5. Concluding remarks and future recommendations are then reserved for Chapter 6.

2. The Language of Thermodynamics

The historical development of thermodynamics has been paved by many people throughout history; through the process of trial and error, we have come to accept certain observations as being *universal* (always true). These observations have resulted in a framework of abstract symbols that can playfully be dubbed "the language of thermodynamics". An important step is to be able to use this common language to interpret, relate, and predict the behaviour of the world around us whilst obeying these universal-*truths*. In practice, these truths are incorporated, amongst many other applications, into external models that can accurately represent the real behaviour of mixtures. Before some of these models are described in detail, however, a brief fundamental review is provided in the present chapter.

2.1. Historical Context

Much of the foundations can be traced back to the necessity of coal mining, which was an economic alternative to society's previous major source of "heat" in the 1800's (wood). As the readily accessible deposits became depleted, surface quarries eventually were developed into mine shafts. As these veins were pursued deeper underground (below water tables) flooding became an obstacle. It was this problem that prompted the re-invention of the steam engine (Falk, 1985; Rarey, 2006; Müller, 2007).

Although the first generation of these engines was highly inefficient (< 5%) they were viewed as a necessity of the *Civilized-World.* Their operating costs were accepted at the time, but as these machines were applied to other industries, serious attention was placed on improving their design (Srinivasan, 2001; Müller, 2007):

> ***"Everyone knows that heat can produce motion. That it possesses vast motive power no one can doubt, in these days when the steam engine is everywhere well known... The study of these engines is of great interest, their importance enormous, their use is continually increasing, and they seem destined to produce a great revolution in the civilized world." (Carnot, 1890)***

2.2. The Fundamental Foundation

Thinking for efficiencies drove people to study the interrelations of heat (Q), work (W), and the *abstract* concept called energy (E). Eventually patterns were identified regarding these variables, and the interrelationships were incorporated into the fundamental theories forming the foundations of thermodynamics.

2.2.1. First Law of Thermodynamics

As touched upon earlier, our collective knowledge has resulted in the identification of certain truths that we consider universal. One such universal-truth is the *conservation of energy*; stated verbatim,

1. Energy can be stored
   Energy can be moved between matter
   Energy can be transformed

These simple statements constitute what is known as the *First Law of Thermodynamics*, and can be mathematically written as

$$\mathrm{E}_{\mathrm{stored}} = \mathrm{E}_{\mathrm{in}} - \mathrm{E}_{\mathrm{out}} \tag{2-1}$$

The concept of *stored* energy $(\mathrm{E}_{\mathrm{stored}})$ naturally lends itself to the characteristic expression "internal energy" (U). However in order to calculate this new quantity, some knowledge of the forms of energy that may enter and/or leave the system is required (where *system* refers to some part of the physical world, separated by a conceptual boundary).

Here I take some liberties on the historical development, and simply state that heat and work have been identified as forms of energy.[3] Where *heat* can be considered transient-energy resulting from temperature differences and *work* can be considered a characteristic-form of energy (typically associated with expenditure). These two energy terms may be used to rewrite the 1st Law in its *traditional form*, a closed system which neglects mass transfer across the system boundaries:

$$\underbrace{d(nu) = dU = \delta Q + \delta W}_{Traditional\ form} \tag{2-2}$$

However this equation cannot be regarded as giving an explicit definition of internal energy, in fact no such definition is known to exist. The postulated existence of internal energy as a property of the system's state (a *state-property*) has proven time and time again to be *consistent* with the 1st Law of Thermodynamics; therefore experience proves its existence:

> One such historical test was the application of the 1st Law to nuclear reactions. For a while it was thought that the 1st Law had a "mass defect" (Müller, 2007), but Einstein was able to establish a relationship between mass and energy $(E = mc^2)$ which further bolstered the 1st Law as a universal-truth. This minor digression makes an important point: an essential part of utilizing the thermodynamic framework, in a consistent manner, lies in the ability to *strictly define* the system being considered.

In accordance, considering a closed system going through a quasi-static change of state (series of very small equilibrium steps where all energy is recovered—considered completely *reversible*),

---

[3] For the interested reader Ingo Müller (2007) has written a good book on the historical development of thermodynamics; although some have criticised his depiction, it nonetheless acts as a good springboard for further indulgence.

where only expansion or compression of the system and heat transfer can take place, enables the following definitions to be made:

$$\begin{aligned}\delta W_{rev} &= -Pd(nv) \\ &= -PdV\end{aligned} \tag{2-3}$$

$$\begin{aligned}\delta Q_{rev} &= Td(ns) \\ &= TdS\end{aligned} \tag{2-4}$$

The total system volume ($V = nv$) and total system entropy ($S = ns$) are considered *extensive* state-properties of the system (i.e. dependent on the mass and state of the system, just like internal energy), and can be used to facilitate the calculation of the change in internal energy by different processes using appropriate substitutions:

$$dU = TdS - PdV \tag{2-5}$$

Since the internal energy is now in a form that *only* depends on the state of the system (i.e. not on the process or path that produces the state) the relationship is also suitable to describe irreversible systems (i.e. physical reality).

2.2.2. Second Law of Thermodynamics

As with the conservation of energy, further historical observations have led to an additional universal-truth; specifically, that energy cannot be transformed or moved without wasting some in the process. This is in reference to the extensive state-property called entropy, a convenience-variable that was introduced by Clausius to represent the quantity $\delta Q/T$; conceptually entropy represents the *wasted effort* (or *lost* energy) involved whenever energy is moved and/or transformed. This is summed up nicely by the following citation (Müller, 2007 p. 71):

> *Clausius summarized his work in the triumphant slogan*
>
> ***Die Energie der Welt ist konstant.***
>
> ***Die Entropie der Welt strebt einem Maximum zu.***
>
> *Die Welt (the universe) was chosen in this statement as being the ultimate thermodynamic system, which presumably is not subject to heating and working, so that $dU = 0$ holds, as well as $dS > 0$.*

This is known as the *Second Law of Thermodynamics*, and is commonly written symbolically using the definition of entropy as

$$dS \geq \frac{\delta Q}{T} \geq 0 \tag{2-6}$$

where the equal sign signifies the limiting value of zero (reversibility). This expression is often combined with the 1st law to obtain the *practical form,*

$$\underbrace{dU \leq TdS - PdV}_{Practical\ form} \qquad (2\text{-}7)$$

The 2nd Law can conveniently be considered as a *constraining relationship* of the 1st Law, which places restrictions on what can and cannot be accomplished in physical reality—nature *naturally* tends towards increasing disorder (entropy).

2.2.3. Third Law of Thermodynamics

The third and last historical observation, like the others, has also proven true 100% of the time; simply stated, that it is impossible to remove all of the heat from an object. This constitutes as the *Third Law of Thermodynamics*, and naturally lends itself to the definition of the lowest point on the *thermodynamic temperature scale* (absolute zero, 0 Kelvin).

The statement implies that it is impossible to reach absolute zero, where all of the *heat* (energy) of a system would be removed (creating a perfect crystal). If thought of in terms of the 2nd Law, it is impossible to have a system with zero entropy; in physical reality this minimum is unrealistic and has never been realized—a universal-truth.[4]

2.2.4. Expanding Results to Open Systems

In the previous subchapters the *traditional* and *practical forms* of the 1st Law were derived on the basis of a *closed system* (no mass transfer across system boundaries); however a form that is *open* to the environment (everything outside the system's conceptual boundary) is required for many practical situations (multiple phases, where each phase is considered a system). For a closed system internal energy was found to be a function of the extensive properties entropy and volume, and can therefore be represented by the total derivative of the continuous-function. The total internal energy of the system, therefore, can be written as the following:

$$nu\ (\text{or } U) = f(ns, nv) = f(S, V) \qquad (2\text{-}8)$$

$$\underbrace{dU = \underbrace{\left[\frac{\partial U}{\partial S}\right]_V}_{T} dS + \underbrace{\left[\frac{\partial U}{\partial V}\right]_S}_{-P} dV}_{Traditional\ form} \qquad (2\text{-}9)$$

Comparing the total derivative with the traditional form of the 1st Law allows the *intensive* state-properties (independent of system mass—temperature and pressure) to be readily identified with their partial derivative equivalents. From this mathematical perspective (entirely consistent with observations), the internal energy may also be extended to open systems via the *conservation of mass* principle:

---

[4] The current world record for the lowest temperature observed stands at 50 pK (NASA, 2003); this is astounding considering that the universe imparts 3 K to bodies through its *background radiation* (Müller, 2007).

$$Conservation\ of\ Mass \Longrightarrow n = n_1 + n_2 + \cdots + n_{Comp} \tag{2-10}$$

$$U = f\left(S, V, n_1, n_2, \cdots, n_{Comp}\right) \tag{2-11}$$

$$\underbrace{dU \le \underbrace{\left[\frac{\partial U}{\partial S}\right]_{V,n}}_{T} dS + \underbrace{\left[\frac{\partial U}{\partial V}\right]_{S,n}}_{-P} dV + \sum_{i=1}^{Comp} \underbrace{\left[\frac{\partial U}{\partial n_i}\right]_{V,S,n_{j\neq i}}}_{\mu_i} dn_i}_{Functional\ form} \tag{2-12}$$

where the *chemical potential* is defined for convenience as the following:

$$\mu_i = \left[\frac{\partial U}{\partial n_i}\right]_{V,S,n_{j\neq i}} \tag{2-13}$$

Like the temperature and pressure of the system, the chemical potential is independent of the quantity contained within the system (an *intensive* state property).

2.3.Auxiliary Properties

Since no entropy meter is known to exist, researchers have been guided to represent the fundamental law by alternative functions that are more readily determined by practitioners (auxiliary properties). Given the functional dependence of internal energy on the intensive $[T, P, \mu_i]$ and extensive $[nv, ns, n]$ state quantities of a system, new properties may be defined by considering various linear changes of these interrelated-variables on internal energy. Legendre transformations facilitate this process (Alberty, 1997; 2001), and can be used to obtain the equivalent potentials of Enthalpy (H), Helmholtz energy (A), and Gibbs energy (G).

2.3.1. Enthalpy (Energy)

$$nh\ (\text{or } H) \equiv (nu) + P(nv) \equiv U + PV \tag{2-14}$$

$$dH = dU + PdV + VdP \tag{2-15}$$

$$H = f\left(S, P, n_1, n_2, \cdots, n_{Comp}\right) \tag{2-16}$$

$$\overbrace{dH \leq \underbrace{\left[\frac{\partial H}{\partial S}\right]_{P,n}}_{T} dS + \underbrace{\left[\frac{\partial H}{\partial P}\right]_{S,n}}_{V} dP + \sum_{i=1}^{Comp} \underbrace{\left[\frac{\partial H}{\partial n_i}\right]_{S,P,n_{j\neq i}}}_{\mu_i} dn_i}^{Functional\ form} \tag{2-17}$$

2.3.2. Helmholtz Energy

$$na\ (\text{or } A) \equiv (nu) - T(ns) = U - TS \tag{2-18}$$

$$dA = dU - TdS - SdT \tag{2-19}$$

$$A = f\left(T, V, n_1, n_2, \cdots, n_{Comp}\right) \tag{2-20}$$

$$\overbrace{dA \leq \underbrace{\left[\frac{\partial A}{\partial T}\right]_{V,n}}_{-S} dT + \underbrace{\left[\frac{\partial A}{\partial V}\right]_{T,n}}_{-P} dV + \sum_{i=1}^{Comp} \underbrace{\left[\frac{\partial A}{\partial n_i}\right]_{T,V,n_{j\neq i}}}_{\mu_i} dn_i}^{Functional\ form} \tag{2-21}$$

2.3.3. Gibbs Energy

$$ng\ (\text{or } G) \equiv \overbrace{(nu) + P(nv) - T(ns)}^{\substack{(nh)-T(ns)\\ P(nv)+(na)}} \equiv \overbrace{U + PV - TS}^{\substack{H-TS\\ PV+A}} \tag{2-22}$$

$$dG = \overbrace{dU + PdV + VdP - TdS - SdT}^{\substack{=dH-TdS-SdT\\ =PdV+VdP+dA}} \tag{2-23}$$

$$G = f\big(T, P, n_1, n_2, \cdots, n_{Comp}\big) \tag{2-24}$$

$$\overbrace{dG \le \underbrace{\left[\frac{\partial G}{\partial T}\right]_{P,n}}_{-S} dT + \underbrace{\left[\frac{\partial G}{\partial P}\right]_{T,n}}_{V} dP + \sum_{i=1}^{Comp} \underbrace{\left[\frac{\partial G}{\partial n_i}\right]_{T,P,n_{j\neq i}}}_{\mu_i} dn_i}^{Functional\ form} \tag{2-25}$$

2.4.Applied Framework

Given the basis upon which the fundamental equations were founded, we have come to believe that fluid properties of homogenous fluids at equilibrium are functions of temperature, pressure, and composition only (Van Ness, et al., 1982). These observations were formalized by Willard Gibbs; for his numerous contributions to the field, it was in his honour that the most readily applied auxiliary function was named. Building atop the fundamental foundations, Gibbs proved that uniformity of temperature, pressure, and composition between the various phases is a necessary criterion to establish an "equilibrium state:"

$$\begin{gathered} dT = 0\ \mathrm{K} \\ dP = 0\ \mathrm{kPa} \\ \sum_{i=1}^{Comp} \mu_i dn_i = 0 \end{gathered} \tag{2-26}$$

Furthermore in order for this condition to be valid over multiple phases, an additional and necessary criterion must be established. Given the framework that has already been constructed, the following must also be true for all components:

$$\sum_{p=1}^{Phase} \mu_i^p dn_i^p = 0 \tag{2-27}$$

In context of the conservation of mass principle, this explicitly requires that the chemical potentials of each component across all phases *must be equal*.

2.4.1. Partial Molar Properties

Given the functional dependence of ng $(\mathrm{or}\ G)$ that has been established, the mathematical definition of exactness can be used to obtain

$$dG = n\left[\frac{\partial g}{\partial T}\right]_{P,n} dT + n\left[\frac{\partial g}{\partial P}\right]_{T,n} dP + \sum_{i=1}^{Comp}\left[\frac{\partial G}{\partial n_i}\right]_{T,P,n_{j\neq i}} dn_i \tag{2-28}$$

Here the chemical potential can formally be defined as (for convenience),

$$\mu_i \equiv \left[\frac{\partial G}{\partial n_i}\right]_{T,P,n_{j\neq i}} \equiv \left[\frac{\partial A}{\partial n_i}\right]_{T,nV,n_{j\neq i}} \equiv \left[\frac{\partial H}{\partial n_i}\right]_{nS,P,n_{j\neq i}} \equiv \left[\frac{\partial U}{\partial n_i}\right]_{nV,nS,n_{j\neq i}} \tag{2-29}$$

Some insight into the nature of the chemical potential can be gained by application of Euler's theorem on homogeneous functions, in this case, of degree zero:

$$G = \sum_{i=1}^{Comp} x_i \mu_i \tag{2-30}$$

This explicitly states that the chemical potential must be a function of temperature, pressure, and composition (as is the Gibbs energy); furthermore, the chemical potentials must also be intensive properties of the system.

It should now be apparent that this quantity is important, but its current form is not entirely useful; as stated elsewhere (Prausnitz, et al., 1986), much of the present work in phase-equilibrium thermodynamics is to relate the *abstract* nature of the chemical potential to physically measurable quantities such as temperature, pressure, and composition. This need becomes especially apparent after integrating the differential forms of H, A, and G (see Chapter 2.3):

$$H = TS + \sum_{i=1}^{Comp} n_i \mu_i \tag{2-31}$$

$$A = -PV + \sum_{i=1}^{Comp} n_i \mu_i \tag{2-32}$$

$$G = \sum_{i=1}^{Comp} n_i \mu_i \tag{2-33}$$

So the Gibbs energy is the only auxiliary property that can be *entirely* related through the component contributions within the system—useful indeed.

2.4.1.1. *Gibbs-Duhem Relation*

In the previous subchapter, the astute observer would have noticed *two* different relationships describing the Gibbs energy, one describing a *change* (Equation (2-28)) and the other an *equilibrium-state* (Equation (2-30)). Moreover each of these relationships was obtained by different mathematical operations, but are they both thermodynamically consistent within the framework that has been built?

In order for both expressions to be correct, they must prove equivalent. If we consider a homogeneous equilibrium state, then any differential change resulting from changes in T, P, n_i must be given by the total differential of G (energy and matter is *conserved*):

$$dG = d\left(\sum_{i=1}^{Comp} x_i \mu_i\right) = \sum_{i=1}^{Comp} x_i d\mu_i + \sum_{i=1}^{Comp} \mu_i dx_i \tag{2-34}$$

For consistency, this expression *must be equivalent* to the exact differential describing the interrelationships of G, T, P, and composition:

$$dG = \left[\frac{\partial G}{\partial T}\right]_{P,x} dT + \left[\frac{\partial G}{\partial P}\right]_{T,x} dP + \sum_{i=1}^{Comp} \underbrace{\left[\frac{\partial G}{\partial n_i}\right]_{T,P,n_{j\neq i}}}_{\mu_i} dx_i \tag{2-35}$$

Accordingly, the following identity is required in order to avoid any contradictions to the thermodynamic framework (built upon observations that have proven universal-truths)—resulting from equating Equation (2-34) and Equation (2-35):

$$\left[\frac{\partial G}{\partial T}\right]_{P,x} dT + \left[\frac{\partial G}{\partial P}\right]_{T,x} dP - \sum_{i=1}^{Comp} x_i d\mu_i = 0 \tag{2-36}$$

This identity is commonly known as the *Gibbs-Duhem equation* (also called the zero-function), and represents a condition that must be obeyed (if you will, a grammatical-law of the language of thermodynamics). Therefore, if a system at constant temperature and pressure is considered, we once again obtain a relationship explicitly stating the concentration dependence of the chemical potential (clearly important). Moreover, it places a restriction on the simultaneous behaviour of T, P, and μ_i for a single phase ($\text{number of components} + 1$ *degree of freedom*).

2.4.2. Fugacity as a Solution Property

Given that temperature and pressure are easily measured, a way to determine the chemical potentials is needed before the framework can be put to practical use. As is often the case with abstract concepts, it is often easiest to generalize from simplified ideal behaviour.

Lewis was the first to consider the chemical potential for a pure ideal gas, and then generalized the results to all systems (Prausnitz, et al., 1986). Using the Gibbs energy function as a starting point, and applying the relationship to one mole of a single-phase pure fluid at isothermal conditions, he obtained

$$dG_i = V_i dP \tag{2-37}$$

Since an ideal gas represents the simplest *known* condition/relationship, the component volume was replaced with the ideal gas equation:

$$dG_i = \frac{RT}{P} dP = RTd(ln\, P) \tag{2-38}$$

The resulting expression quantifies the relationship between the Gibbs energy and pressure for an ideal gas at constant temperature. To generalize his result Lewis introduced a new function *fugacity* to stand for the *true* (observable) system pressure, compensated for by molecular interactions. This is written as,

$$dG_i = RT \overbrace{d(ln\, f)}^{\substack{Ideal\ gas \\ correction}} \tag{2-39}$$

where the equation is then integrated from a *known* standard state (designated by a superscript 0) to obtain the following relationship:

$$G - G^0 = RTln\frac{f}{f^0} \tag{2-40}$$

In order for the fugacity definition to be universally valid, however, the relationship must reduce to the ideal gas value at the pressure limit (the basis of the derivation). Thus, the following criterion is a *necessary* component of the fugacity definition:

$$\lim_{P\to 0}\left(\frac{f}{P}\right) = 1 \tag{2-41}$$

2.4.2.1. *An Additional Criterion for Phase Equilibrium*

As mentioned earlier, the equality of the chemical potential of a component in all phases is a criterion for phase equilibrium. Given the abstract nature of chemical potential a new property was defined for convenience – fugacity. Using the fundamental framework, the fugacity will be shown to be a suitable alternative to chemical potential in describing the state of equilibrium. For each component (i) in solution, the fugacity (conveniently thought of as a *utility-function*) can be rewritten as

$$d\mu_i \equiv d\bar{G}_i = RTd\ln \hat{f}_i \tag{2-42}$$

$$\therefore \mu_i = \bar{G}_i = G_i^0 + RT\ln\frac{\hat{f}_i}{f_i^0} \tag{2-43}$$

$$\lim_{P\to 0}\left(\frac{\hat{f}_i}{x_i P}\right) = 1 \tag{2-44}$$

where the circumflex (^) distinguishes the component-fugacity from the solution property (f). For each component in each phase (p); integration at constant temperature yields

$$\mu_i^p(Final) = \mu_i^p(Initial) + RTd\ln\frac{\hat{f}_i^p(Final)}{\hat{f}_i^p(Initial)} \tag{2-45}$$

Since the integrand is typically evaluated from a known standard state, the *same* initial state can be selected for each component in each phase of the system (it is completely arbitrary). Given the equality of chemical potentials at equilibrium, it can be easily shown that the fugacities in each phase must also have the same value. For a hypothetical system consisting of phases α through π it can be easily shown that

$$\hat{f}_i^\alpha(Final) = \hat{f}_i^\beta(Final) = \cdots = \hat{f}_i^\pi(Final) \tag{2-46}$$

This constitutes a major justification for the introduction of the fugacity as a thermodynamic variable (Van Ness, et al., 1982).

2.4.2.2. *Ideal Mixture*

Given the usefulness of fugacity in representing the chemical potential, we can properly define an *ideal mixture* by integrating the utility-function from the *pure* state to the actual state of the *mixture*:

$$\mu_i - \mu_i^0 = RT\ln\frac{\hat{f}_i^{is}}{f_i} = RT\ln\frac{x_i f_i}{f_i} = RT\ln x_i \tag{2-47}$$

Keeping in mind that the chemical potential is an intensive property that, like the Gibbs energy, depends on temperature, pressure, and composition, an *ideal mixture* can be formally defined as

$$\therefore \mu_i(P,T,x) \equiv \mu_i^0(P,T) + RT\ln x_i \tag{2-48}$$

This definition will prove to be useful in developing the following chapter.

2.4.3. Excess Properties & Activity Coefficients

Since many of the fundamental properties we wish to use are in the form of derivatives, it is often helpful to adopt a standard datum from which to calculate the change. For this purpose it is common practice to relate properties to their *excess* values – property deviations from their ideal-solution values (i.e. from a known condition to an actual condition). This concept can be generalized as

$$M^E = M - M^{is} \tag{2-49}$$

where M represents any *real/actual* property value, M^{is} the ideal solution value, and M^E the excess quantity of that property value; since this definition is only shifting the original value, all of the preceding energy relationships (including partial derivatives) can be directly written in terms of these excess values. By applying this definition the chemical potential may be written as,

$$\mu_i^E = \mu - \mu^{is} = \underbrace{[\mu - \mu_i]}_{RT\ln\frac{f}{f_i}} - RT\ln x_i \tag{2-50}$$

Given that this relationship has proven quite useful in practice, it is common to introduce a convenience variable known as the *activity coefficient:*

$$\gamma_i \equiv \frac{\hat{f}_i}{x f_i} \tag{2-51}$$

$$\therefore \mu^E = RT\ln\gamma_i \equiv \left[\frac{\partial(G^E/RT)}{\partial n_i}\right]_{T,P,n_{j\neq i}} \tag{2-52}$$

This means that for a real mixture, the chemical potential can be represented by an augmented ideal mixture value:

$$\mu_i(P,T,x) = \mu_i^0(P,T) + RT\ln x_i + \mu^E \tag{2-53}$$

$$\mu_i(P,T,x) = \mu_i^0(P,T) + RT\ln x_i + \underbrace{RT\ln\gamma_i\,(T,P,x)}_{\substack{Excess\ contribution\\ to\ the\ ideal\ solution\\ value}} \tag{2-54}$$

2.5.Lattice Theory

Many of the currently used methods are based, at least in some part, upon Lattice Theory. Here the liquid state is viewed as quasi-crystalline state where the individual molecules are *confined* to a region of space within the system (called a *lattice*). The goal of the theory is to determine the energy of the liquid—in classical thermodynamics this is the internal energy U—pure or mixed, by accounting for the energies (E_i) of each probable state (p_i^*) of the mixture:

$$U = E = \sum_{i}^{States} p_i^* E_i \tag{2-55}$$

The last equation can be written in terms of the *canonical partition function,* where the system is allowed to exchange heat with the environment at fixed temperature, volume, and number of particles. For a binary mixture the equation takes the form (Prausnitz, et al., 1986),

$$Q = \sum_{N12} g(N_1, N_2, N_{12}) exp\left[-\frac{E_{total}}{kT}\right] \tag{2-56}$$

where $g(N_1, N_2, N_{12})$ represents the number of ways molecules of component 1 and 2 can be arranged on the lattice, and the latter term represents the total energy of the lattice.

Theoretical representation of fluid properties is still one of the most complicated and challenging tasks of chemical physics. An exact solution to the problem would necessitate answering the following questions (Malanowski, et al., 1992):

1. What are the intermolecular forces between molecules?

   This falls into the realm of quantum mechanics and can be solved exactly only for very simple cases.
2. What is the effect of intermolecular interactions on the spatial arrangement of molecules and, subsequently, their thermodynamic properties?

   In principle this can be answered by statistical mechanics if a solution to the first problem is known.

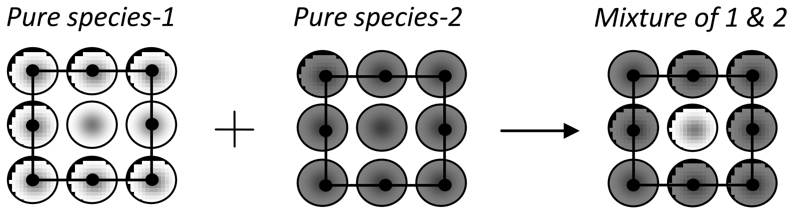
Given that no universal theory has yet been found that meets the needs of industry, methods such as Wilson, NRTL, UNIQUAC are often used. Methods like these tend to include a tuning parameter to account for the intermolecular forces between molecules, along with some kind of assumption on the arrangement of the molecules (e.g. local composition & Guggenheim's quasichemical approximation).

2.5.1. Quasichemical Approximation

Guggenheim constructed a lattice theory for molecules of equal size which form mixtures that are not necessarily random ($s^E \neq 0$). Assumptions on the spatial arrangement of molecules within the mixture were made, which resulted in a simplistic representation of mixtures known as the *quasichemical approximation*; given that the derivation is immaterial for discussion purposes the result will just be simply stated as the following (Prausnitz, et al., 1986):

$$\frac{g^E}{RT} = \left(\frac{w}{kT}\right) x_1 x_2 \left[1 - \frac{1}{2}\left(\frac{2w}{zkT}\right) x_1 x_2 + \cdots\right] \tag{2-57}$$

where w represents the *interchange energy* from each new 1-2 contact formed when pure species-1 is mixed with pure species-2, and z represents the number of nearest touching molecules (called the *coordination number*)—see Figure 1.



Where in this figure, $z = 8$

Pure species-1 (entirely made up of 1-1 contacts, each having potential energy ε_{11}) is mixed with pure species-2 (entirely made up of 2-2 contacts, each having potential energy ε_{22}) to form a mixture containing 1-2 contacts (each having potential energy ε_{12}) by replacing 1-1 and 2-2 interactions:

$$(1-1) + (2-2) \rightarrow 2(1-2)$$

$$\varepsilon_{11} + \varepsilon_{22} \rightarrow 2\varepsilon_{12}$$

$$\Delta\varepsilon_{12} = \varepsilon_{12} - \frac{1}{2}(\varepsilon_{11} + \varepsilon_{22})$$

Figure 1 Physical significance of interchange energy w, where the energy absorbed in the above illustration is 2w (Prausnitz, et al., 1986).

where the assumption of nonrandomness can be written as (Renon, et al., 1968),

$$x_{21} x_{12} = (1 - x_{21})(1 - x_{12}) exp\left[-\frac{1}{z}(2W_{12} - W_{11} - W_{22})/RT\right] \tag{2-58}$$

Therefore, when two components are mixed, the molecules of these components are not completely random. Intuition also tells us this, because if it is assumed that these molecules experience repulsion and attraction forces, then some spatial orientation (structure or *nonrandomness*) must result to minimize the resistances of the intermolecular forces involved between the mixture molecules (unless these forces are the exact same for every molecular combination). This concept has been employed in the development of the NRTL and UNIQUAC equations, which will be discussed in further detail in the following literature review chapter.

3. LITERATURE REVIEW

Given that the main goal of this work is to develop a model that will be of use to industry, an understanding of models currently in use is in order. All proposed equations (no matter the form) are intended to "stand-in" for the Gibbs excess energy relationship

$$g^E = u^E - Ts^E + Pv^E = h^E - Ts^E \tag{3-1}$$

where by definition, the following boundary conditions must be obeyed:

$$\begin{aligned} g^E &\longrightarrow 0 \\ x_i &\longrightarrow 0 \end{aligned} \tag{3-2}$$

The simplest assumption one can make about g^E is to set it equal to zero (ideal solution) – industrial mixtures, however, are rarely ideal. The next simplest assumption obeying these boundary conditions is the so-called Porter equation (Porter, 1921):

$$g^E = Ax_1x_2 \tag{3-3}$$

In this equation, A is an adjustable parameter which can be fitted to experimental data. Additional simplifications can also be made by setting either s^E or h^E equal to zero and then developing a relationship to describe the remaining property/variable.

3.1.Algebraic Expressions

If it is assumed that s^E = 0 (a concept of regular solutions), then a relationship for h^E that could be solved numerically is needed in order to obtain g^E. Historically h^E was represented by polynomial expansions in the mole fractions or volume fractions, like those resulting in the Margules and Redlich-Kister (1948) expressions.

A general problem with these expansions is that the g^E of a mixture is split into terms corresponding to binary, ternary and higher systems formed by *all* possible components present in the mixture (Malanowski, et al., 1992):

$$g^E = \sum_{i=1}^{n}\sum_{j=i+1}^{n} g_{i,j}^E + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n} g_{i,j,k}^E + \cdots \tag{3-4}$$

This relation would require equations describing binary, ternary and higher-order interactions. Such a description is highly unrealistic, because the number of experimental data required for the fitting of model parameters increases for each term with the second power of the number of components in the mixture. As with most things simplifications are often prudent, and have led to some historically, successful relationships (e.g. truncated version of Margules equation).

3.1.1. Pros & Cons

Polynomial expansions were the first models to provide a reasonable representation of *real* mixtures (adhered to the Gibbs-Duhem equation). Many of them can be used today with equal success, but with limitations compared to currently accepted methods. Some observations of their use are listed below for convenience:

- ☑ They are easily applied to two-component systems, and their "tuning" parameters tend to have some limited physical significance.
- ☒ They usually cannot be extended to predict ternary and higher mixtures from binary parameters (except for very simple mixtures).
- ☒ Terms can be added as needed and parameters can be defined as functions of temperature to obtain very accurate representation of g^E or phase equilibrium data over a wide temperature range.
  - ➲ This, however, requires a sufficient amount of experimental information to regress the higher-order interactions of the expanded equation, and thus contributes to the limited applicability of such equations—they are not easily extended to mixtures of more than two components.
- ☑ The equations provide the most flexible description of g^E.
  - ➲ In view of this, the methods have found some modern applications in smoothing experimental excess function data (Byer, et al., 1973; Malanowski, et al., 1992; Van Ness, 1995).
- ☑ They are often given the merit of being mathematically simple to implement.
  - ➲ This, however, is now of little importance due to the wide availability of computers.

3.2. Local Composition Methods

An alternate starting point is to assume that h^E = 0; this leads to the concept of athermal solutions which has been developed by Flory, Huggins, and others in the study of polymer solutions (Orye, et al., 1965). The Flory-Huggins relationship was successfully modified by Wilson, and has spawned a group of successful models known as *local composition methods*.

3.2.1. Wilson Equation

Although not stated directly in Wilson's original paper, his development can essentially be considered as an ad-hoc two-fluid-extension of a modified Flory-Huggins relationship. He considered mixtures of molecules which not only differ in size, but which also differ in their intermolecular forces (Prausnitz, et al., 1986). In this case the Flory-Huggins equation for athermal mixtures is given by

$$\frac{g^E}{RT} = \underbrace{\sum_{i=1}^{Comp} x_i \ln x_i}_{\text{ideal solution case}} \xrightarrow{\text{and is replaced by}} \sum_{i=1}^{Comp} x_i \ln \frac{\Phi_\mathrm{i}}{x_i} \tag{3-1}$$

where Φ_i is the volume fraction and x_i is the mole fraction of the *i*th molecule, and provides a simple measure of a system's asymmetry; this size affect is given by,

$$\Phi_{\mathrm{i}} = \frac{x_i \upsilon_i^L}{\sum_{j=1}^{N} x_j \upsilon_j^L} \tag{3-2}$$

Instead of directly weighting the component sizes to the solution composition, Wilson simply replaced $\ln x_i$ with what he termed the "local composition:"

$$\frac{x_{ij}}{x_{ii}} = \frac{x_j e^{-\varepsilon_{ij}/RT}}{x_i e^{-\varepsilon_{ii}/RT}} \tag{3-3}$$

so that the ratio of i and j around a central i molecule (x_{ii} and x_{ij} respectively) is related to their contact energies (ε_{ij} and ε_{ii})—the probability of finding i or j as a neighbour. Wilson used this definition to redefine the volume fractions into the *local* volume fractions ξ_i and ξ_j (Orye, et al., 1965):

$$\frac{g^E}{RT} = \underbrace{\sum_{i=1}^{Comp} x_i \ln x_i}_{\text{ideal solution case}} \xrightarrow{\text{replaced by}} \sum_{i=1}^{Comp} x_i \ln \frac{\xi_i}{x_i} \tag{3-4}$$

$$\xi_i = \frac{x_i \upsilon_i^L e^{-\varepsilon_{ij}/RT}}{\sum_{j=1}^{N} x_j \upsilon_j^L e^{-\varepsilon_{ij}/RT}} \tag{3-5}$$

The g^E-form of the Wilson equation is then obtained by substituting the local volume fractions into the Flory-Huggins equation, and then introducing the convenience parameter Λ_{ij}:

$$\frac{g^E}{RT} = -\sum_{i=1}^{Comp} x_i ln\left[\sum_{j=1}^{Comp} x_j \Lambda_{\mathrm{ij}}\right] \tag{3-6}$$

where

$$\Lambda_{\mathrm{ij}} = \frac{\upsilon_j^L}{\upsilon_i^L} e^{-(\varepsilon_{ij}-\varepsilon_{ii})/RT} \tag{3-7}$$

So instead of directly accounting the *primary* energies of each type of contact (i.e. $\varepsilon_{ij} \equiv \varepsilon_{ji}$, ε_{ii}, and ε_{jj}) he simplified by introducing *secondary* interaction parameters Λ_{ij} and Λ_{ji}. This enabled Wilson to reduce the number of parameters required for each binary-pair by one (Marcilla, et al., 1995).

From the Wilson equation the following expression for the activity coefficients can be derived (using Equation (2-52)):

$$\ln \gamma_i = -\ln\left(\sum_{j=1}^{Comp} x_j \Lambda_{\mathrm{ij}}\right) + 1 - \sum_{j=1}^{Comp}\left(\frac{x_j \Lambda_{\mathrm{ji}}}{\sum_{k=1}^{Comp} x_k \Lambda_{jk}}\right) \tag{3-8}$$

So like the Flory-Huggins relationship, the Wilson equation contains no explicit part to account for the enthalpic effects of mixing; however unlike the Flory-Huggins equation, the entropic contribution to the Gibbs energy is accounted for by the local composition concept which requires knowledge about the energetic interactions between the mixture molecules.

3.2.1.1. *Pros & Cons*

- ☑ As can be seen by the final equation form, this model is applicable to mixtures of more than two components, and only requires binary parameters relating to the binary interactions between component-pairs.
- ☑ Molecular size differences are taken into account by the ratio of pure component volumes (essentially weighting the binary interaction contributions according to the size of the components), and has been proven to represent both polar and nonpolar mixtures quite well (Orye, et al., 1965; Bruin, 1970; Walas, 1985; Prausnitz, et al., 1986; Malanowski, et al., 1992; Poling, et al., 2001).
- ☑ When the parameters $(\varepsilon_{ij} - \varepsilon_{ii})$ are treated as being temperature independent, the equation can only be applied precisely for isothermal data (Malanowski, et al., 1992).
  - ➲ However mild temperature dependence is built into the equation via $e^{-(\varepsilon_{ij}-\varepsilon_{ii})/RT}$, and usage over a limited temperature range is often possible. For a reliable representation of phase equilibria *and* excess properties, however, parameters explicit in temperature should be used.
- ☒ The Wilson equation, as given, cannot handle liquid-liquid immiscibility.
  - ➲ Wilson (1964; Scatchard, et al., 1964) demonstrated that the g^E prediction could be modified by including an arbitrary multiplier C (later shown to be derivable from the two-liquid theory used to formulate NRTL— (Renon, et al., 1969)):

$$\frac{g^E}{RT} = -C \sum_{i=1}^{Comp} x_i ln \left[\sum_{j=1}^{Comp} x_j \Lambda_{\mathrm{ij}} \right] \tag{3-9}$$

Unfortunately this modification is not easily extended to multicomponent systems; as the tuning-value would be unique for each binary.

3.2.2. NRTL Equation

Renon & Prausnitz (1968) introduced the concept of the two-fluid model and combined it with Wilson’s local composition concept; inclusion of a *nonrandomness* correction laid the foundation for the development of the NRTL equation (non-random, two-liquid).

Scott (1956) proposed that instead of treating mixtures as single-fluids (effectively *averaging* the solution environment), a mixture could be considered as a solution of two *distinct* fluids (one with molecule i at the centre, and one with molecule j at the centre). Renon adopted this strategy and expressed g^E in terms of two hypothetical fluids, enabling g^E to be calculated *directly* via

$$g^E = x_i \left(g^{(i)} - g^{(i)}_{pure} \right) + x_j \left(g^{(j)} - g^{(j)}_{pure} \right) \tag{3-10}$$

where $\left(g^{(i)} - g^{(i)}_{pure}\right)$ represents the energy of transferring a pure species into a *hypothetical* mixture environment with energy $g^{(i)}$. In order to account for the number of *j-i*, *i-i*, *i-j*, and *j-j* contacts, this strategy was combined with the local composition construct (the mean local composition around the central i and j molecules). The mixture environments were redefined in terms of local mole fractions as

$$\frac{x_{ji}}{x_{ii}} = \frac{x_j}{x_i}\left[\frac{e^{-\alpha g_{ji}/RT}}{e^{-\alpha g_{ii}/RT}}\right] \qquad (3\text{-}11)$$

where Renon introduced what he called a *nonrandomness parameter* to scale down the characteristic energy differences between j around i and i around i (and vice versa, where $g_{ji} \equiv g_{ij}$); making the appropriate substitutions into Equation (3-11) (see Figure 2), the expression can be rewritten as

$$x_{ji} = \frac{x_j exp\left[-\alpha\left(g_{ji} - g_{ii}\right)/RT\right]}{x_j + x_i exp\left[-\alpha\left(g_{ij} - g_{jj}\right)/RT\right]} \qquad (3\text{-}12)$$

$$g^E = \sum_{i=1}^{Comp} x_i x_{ji}\left(g_{ji} - g_{ii}\right) \qquad (3\text{-}13)$$

Upon the appropriate differentiation, the NRTL expression for activity coefficients is obtained (see Equation (2-52)):

$$\ln\gamma_i = \frac{\sum_{j=1}^{Comp} \tau_{ji}G_{ji}x_j}{\sum_{k=1}^{Comp} G_{ki}x_k} + \sum_{j=1}^{Comp} \frac{x_j G_{ij}}{\sum_{k=1}^{Comp} G_{kj}x_k}\left(\tau_{ij} - \frac{\sum_{k=1}^{Comp} \tau_{kj}G_{kj}x_k}{\sum_{k=1}^{Comp} G_{kj}x_k}\right) \qquad (3\text{-}14)$$

using the convenience-variables,

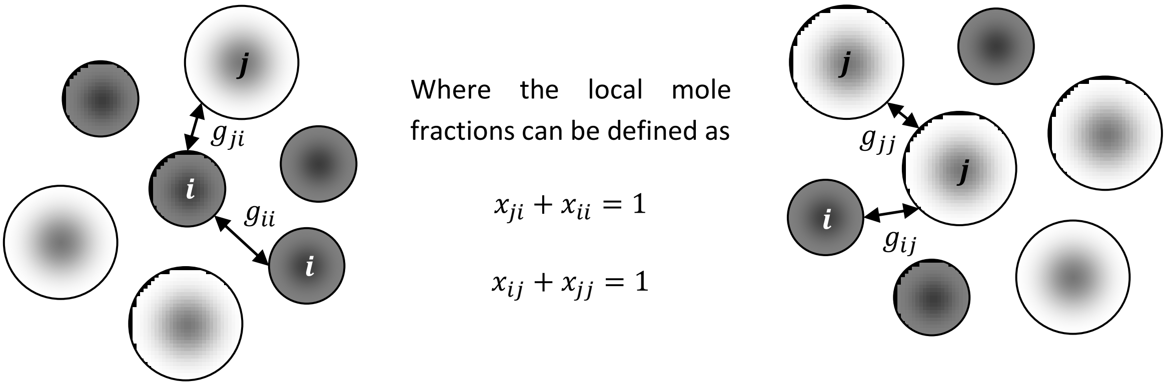
$$\tau_{ji} = \frac{\left(g_{ji} - g_{ii}\right)}{RT} \qquad (3\text{-}15)$$

$$G_{ji} = exp\left(-\alpha\tau_{ji}\right) \qquad (3\text{-}16)$$

These definitions can also be used to redefine the excess Gibbs function of the NRTL equation; expressed for a mixture of more than two components as

$$\frac{g^E}{RT} = \sum_{i=1}^{Comp} x_i x_{ji}\left(g_{ji} - g_{ii}\right) = \sum_{i=1}^{Comp} x_i \frac{\sum_{j=1}^{Comp} \tau_{ji}G_{ji}x_j}{\sum_{k=1}^{Comp} G_{ki}x_k} \qquad (3\text{-}17)$$

So the NRTL equation, unlike the Wilson’s equation, directly calculates the deviation in the Gibbs energy from the ideal Gibbs energy (i.e. it never calculates the entropic part of mixing), and is able to represent mixtures with two liquid phases. As no measure of component size was introduced, however, the equation is limited to *approximately* symmetric mixtures.



Where the local mole fractions can be defined as

$$x_{ji} + x_{ii} = 1$$

$$x_{ij} + x_{jj} = 1$$

Combining the two methodologies together (i.e. local composition & two-liquid theory), the residual Gibbs energy of each hypothetical fluid $\left(g^{(i)}\right)$ can be written as

$$g^{(i)} = x_{ii}g_{ii} + x_{ji}g_{ji}$$

where pure liquid (i) corresponds to the local mole fractions $x_{ii} = 1$ and $x_{ji} = 0$; reducing to

$$g_{pure}^{(i)} = g_{ii}$$

Figure 2 Two cells according to Scott's two-liquid theory of binary mixtures, including the relationships enabling its application to the local composition concept.

3.2.2.1. *Nonrandomness Parameter*

One of the main benefits of the NRTL equation is its ability to represent mixtures exhibiting partial miscibility. Unlike the empirically modified Wilson equation (Wilson, 1964; Scatchard, et al., 1964) the NRTL equation is able to represent mixtures of more than two components if given an appropriate nonrandomness factor.

In the original derivation of the NRTL equation, relationships were drawn between the nonrandomness factor (α) and Guggenheim's quasichemical approximation (see 2.5.1)—it was shown that α was related to the inverse of the coordination number $(1/z)$ that appeared in Guggenheim's expression (which can be *conceptually* considered as $2/z$). Renon made an effort to explicitly mention that the nonrandomness parameter was *entirely empirical*; however, it appears that he was clearly guided by the idea that the two quantities were related.

The defining NRTL relations can be manipulated into a form explicitly showing the nonrandomness as (Renon, et al., 1968):

$$x_{ji}x_{ij} = \left(1 - x_{ji}\right)\left(1 - x_{ij}\right)\exp\left(-\alpha\left(2g_{ij} - g_{ii} - g_{jj}\right)/RT\right) \tag{3-18}$$

Since the coordination number is typically found to be within the range of 6-12 (Renon, et al., 1968; Abrams, et al., 1975; Prausnitz, et al., 1986), the alpha parameter was expected to be on the order 0.1 – 0.3 (later revised to 0.2 – 0.47 based on experimental fittings/experience). He even went so far as to suggest values based on the chemical nature of the mixtures being considered (Renon, et al., 1968)—clearly influenced by the *idea* that alpha was a property of the mixture. Subsequent studies in the field, however, have eliminated any physical significance originally attributed to the parameter (Marina, et al., 1973; Heidemann, et al., 1973; Tassios, 1976).

3.2.2.2. *Pros & Cons*

- ☑ Retains the beneficial attributes exhibited by the Wilson equation, and has the additional advantage of being able to describe LLE with a suitable value for α.
- ☑ The LLE tuning was introduced in a consistent way, thus enabling readily extensions to multicomponent systems.
    - ➲ This modification, unlike the modification originally proposed by Wilson, does not require unique factors for each component-binary in the mixture.
- ☑ Like in the case of Wilson's equation there is mild temperature dependence built-in (τ_{ji}).
    - ➲ Explicit temperature dependence was later formulated where all three parameters of the NRTL equation were treated as linear functions of temperature (Nagata, et al., 1973; Walas, 1985; Malanowski, et al., 1992); however for the simultaneous correlation of VLE, γ^{∞}, H^E, C_p, y_{AZD}, SLE, and LLE data a quadratic dependence is often required (Rarey, 2009-2010). Since explicit temperature dependence increases the number of parameters, it is generally only used on an "as needed basis" in practice.
- ☒ Compared to the original Wilson equation the NRTL model requires the regression of three parameters instead of just two (i.e. conventional wisdom tells us that more experimental data is required to obtain an accurate fit).
    - ➲ Renon tried to address this issue by attempting to *predict* the nonrandomness based on the *mixture-types* being evaluated (Renon, et al., 1968; Renon, et al., 1969; Bruin, et al., 1971), however as mentioned, subsequent studies have eliminated any physical significance originally attributed to the parameter (e.g. Tassios 1976, even showed that $\alpha = -1$ works in many cases).
    - ➲ Further work was conducted in an attempt to reduce the NRTL expression to only one parameter by introducing predictions based on pure component properties and molecular structure (Bruin, et al., 1971; Hála, 1972); however, the simplification leads to a loss in accuracy, as compared to the original form of the NRTL equation.
    - ➲ In practice the nonrandomness factor (α) is unceremoniously set to 0.2 or 0.3 (typically results in *stable* predictions with the NRTL equation), and is typically only treated as an adjustable parameter when warranted (i.e. to fit LLE, or in order to stabilize the prediction)—it is entirely considered an *empirical factor*.
- ☑ One of the major benefits of the NRTL model is its increased flexibility in representing the g^E curve over the entire composition range (Morisue, et al., 1972; Walas, 1985).
    - ➲ However, its flexibility has also led to some instability whereby two two-phase regions are sometimes predicted (Katayama, et al., 1973), hence the range of its intended application should be evaluated for any instability. This instability is often overcome by setting/changing the value of alpha in practice.

3.2.3. UNIQUAC Equation

To combine the advantages of the Wilson and NRTL equations Abrams and Prausnitz (1975) proposed a two-parameter equation that extends the quasi-chemical theory of Guggenheim (Malanowski, et al., 1992)—the universal quasi-chemical (UNIQUAC) equation. Like the NRTL equation, the UNIQUAC equation can be derived from Scott's two-fluid theory; a brief derivation following Maurer and Prausnitz (1978) will be made.

Like the derivation of the NRTL equation (Renon, et al., 1968), it was assumed that the fluid was made up of two distinct fluids based on Scott's two-fluid theory (1956). Abrams and Prausnitz, however, approached this from a statistical mechanics point of view. This required them to actually attempt to count all of the molecular-molecular interactions occurring in a binary mixture. This resulted in an expression in terms of the excess internal energy u^E, instead of the excess Gibbs energy expression used in the NRTL derivation (Equation (3-10)):

$$u^E = \overbrace{\frac{1}{2} z x_1 q_1 N_A \left[\theta_{11} U_{11}^{(1)} + \theta_{21} U_{21}^{(1)} - U_{11}^{pure1}\right]}^{Hypothetical\ fluid\ (1)} + \underbrace{\frac{1}{2} z x_2 q_2 N_A \left[\theta_{22} U_{22}^{(2)} + \theta_{12} U_{12}^{(2)} - U_{22}^{pure2}\right]}_{Hypothetical\ fluid\ (2)} \tag{3-19}$$

where z is the coordination number, q_i is the surface area of the i^{th} component, N_A is Avogadro's number, $U_{21}^{(1)}$ represents the potential energy of 2-1 neighbours in hypothetical fluid (1)—similar explanations for other combinations—and where the local surface fractions (θ) obey the following conservation equation for each component i in the mixture $(i = 1\ to\ Comp)$:

$$\sum_{j=1}^{Comp} \theta_{ji} = 1 \tag{3-20}$$

and similar to Wilson (1964), the *local* surface fractions are defined by (and vice versa for θ_{ij}/θ_{jj}),

$$\frac{\theta_{ji}}{\theta_{ii}} = \frac{\theta_j}{\theta_i} exp\left(-\frac{\frac{1}{2} z\left(U_{ji} - U_{ii}\right)}{kT}\right) \tag{3-21}$$

where the surface area fractions are given by

$$\theta_i = \frac{q_i x_i}{\sum_{\mathrm{j=1}}^{\mathrm{Comp}} q_j x_j} \tag{3-22}$$

$$q_i = \sum_{\mathrm{k=1}}^{\mathrm{Group}} \nu_k^{(i)} Q_k \tag{3-23}$$

where

$\nu_k^{(i)}$: is the number of groups of the k^{th} type in the i^{th} molecule

Q_k : is the group surface area of the k^{th} type in the i^{th} molecule

q_i : is the surface area (shape) parameter of the i^{th} molecule

Given the conservation equations, and assuming that $U_{11}^{(1)} = U_{11}^{pure1}$ and $U_{22}^{(2)} = U_{22}^{pure2}$, Equation (3-19) can then be simplified as

$$u^E = x_1\theta_{21}q_1\underbrace{\left\{\frac{1}{2}zN_A(U_{21} - U_{11})\right\}}_{\Delta u_{21}} + x_2\theta_{12}q_2\underbrace{\left\{\frac{1}{2}zN_A(U_{12} - U_{22})\right\}}_{\Delta u_{12}} \tag{3-24}$$

Although internal energy is more intuitive, an expression for g^E is often needed for practical calculations. The process of translating the expression for u^E into a useful expression for g^E can be broken-up into three steps. The first step involves using one of the subtleties of thermodynamics, where the expression for u^E is translated into an expression for a^E (Elliott, et al., 1999),

$$T\left(\frac{\partial(A/RT)}{\partial T}\right)_V = \frac{T}{RT}\left(\frac{\partial U}{\partial T}\right)_V - \frac{TU}{RT^2} - \frac{T}{R}\underbrace{\left(\frac{\partial S}{\partial T}\right)_V}_{=\frac{1}{T}\left(\frac{\partial U}{\partial T}\right)_V} = -\frac{U}{RT} \tag{3-25}$$

This requires the integration of Equation (3-25) via the substitution of Equation (3-24). Assuming that Δu_{ji} and Δu_{ij} are independent of temperature yields the following:

$$\begin{gathered}\int_{1/T_0}^{1/T} d\left(\frac{a^E}{RT}\right) = \left.\frac{a^E}{RT}\right|_{1/T} - \left.\frac{a^E}{RT}\right|_{1/T_0} = -\int_{1/T_0}^{1/T} d\left(\frac{u^E}{RT}\right)\frac{dT}{T} \\ \Rightarrow \frac{a^E}{RT} = \int_{1/T_0}^{1/T} d\left(\frac{-u^E}{RT^2}\right)dT + C\end{gathered} \tag{3-26}$$

$$\frac{a^E}{RT} = -x_1q_1\ln\left(\theta_1 + \theta_2\exp\left(-\frac{\Delta u_{21}}{RT}\right)\right) - x_2q_2\ln\left(\theta_2 + \theta_1\exp\left(-\frac{\Delta u_{12}}{RT}\right)\right) + C \tag{3-27}$$

where *C* is a constant of integration which can be evaluated at the limit as $1/T_0 \rightarrow 0$. At very high temperatures the enthalpic contribution becomes very small compared to the entropic one, and it can be assumed that components *i* and *j* form an athermal mixture.

The second step involves using the equation of Guggenheim for athermal mixtures (for molecules of arbitrary size and shape) as the boundary condition (Maurer, et al., 1978):

$$\left.\frac{a^E}{RT}\right|_{1/\mathrm{T}_0} = \left(\frac{a^E}{RT}\right)_{athermal} = -\left(\frac{s^E}{R}\right)_{combinatorial}$$
$$= \sum_{i=1}^{Comp}\left(x_i \ln\frac{\Phi_\mathrm{i}}{x_i}\right) + \frac{z}{2}\sum_{i=1}^{Comp}\left(q_i x_i \ln\frac{\theta_\mathrm{i}}{\Phi_\mathrm{i}}\right) \quad (3\text{-}28)$$

where the segment/volume fractions (Φ_i) are calculated using the following definitions:

$$\Phi_\mathrm{i} = \frac{r_i x_i}{\sum_{\mathrm{j=1}}^{\mathrm{Comp}} r_j x_j} \quad (3\text{-}29)$$

$$r_i = \sum_{\mathrm{k=1}}^{\mathrm{Group}} \nu_k^{(i)} R_k \quad (3\text{-}30)$$

where

$\nu_k^{(i)}$: is the number of groups of the *k*th type in the *i*th molecule
R_k : is the group volume of the *k*th type in the *i*th molecule
r_i : is the volume (size) parameter of the *i*th molecule

The third and final step involves using the assumption of Hildebrand and Scott to obtain an expression for g^E, which makes use of the observation that at low pressures the excess volume of mixing (relative to the other contributions) is quite small (Elliott, et al., 1999):

$$(a^E)_{T,V} \approx (g^E)_{T,P} = g^{E(C)} + g^{E(R)} \quad (3\text{-}31)$$

where the *combinatorial* term $\left(g^{E(C)}\right)$ represents the entropic contribution for the case of random mixing (resulting from molecular differences in size and shape), and where the *residual* $\left(g^{E(R)}\right)$ accounts for the enthalpic contributions (energetic interactions of the solution and differences in entropy of mixing between the local composition and the random mixture). With the appropriate partial differentiation, the activity coefficients can be obtained from Equation (2-52):

$$\ln\gamma_i = \ln\gamma_i^C + \ln\gamma_i^R \quad (3\text{-}32)$$

$$\ln\gamma_i^C = \ln\frac{\Phi_\mathrm{i}}{x_i} + \frac{z}{2}q_i\ln\frac{\theta_\mathrm{i}}{\Phi_\mathrm{i}} + l_i - \frac{\Phi_\mathrm{i}}{x_i}\sum_{j=1}^{Comp}\left(x_j l_j\right)$$
$$= \ln\Phi_i' + \frac{z}{2}q_i\ln\frac{\theta_i'}{\Phi_i'} + l_i - \Phi_i'\sum_{j=1}^{Comp}\left(x_j l_j\right) \quad (3\text{-}33)$$

$$\begin{aligned}\ln\gamma_i^R &= q_i\left[1-\ln\left(\sum_{j=1}^{Comp}\theta_j\tau_{ji}\right)-\sum_{j=1}^{Comp}\frac{\theta_j\tau_{ij}}{\sum_{k=1}^{Comp}\theta_k\tau_{kj}}\right]\\ &= q_i\left[1-\ln\left(\sum_{j=1}^{Comp}x_j\theta_j'\tau_{ji}\right)-\sum_{j=1}^{Comp}\frac{x_j\theta_j'\tau_{ij}}{\sum_{k=1}^{Comp}x_k\theta_k'\tau_{kj}}\right]\end{aligned} \tag{3-34}$$

where l_i has been defined for convenience, and where θ_i' and Φ_i' are often introduced to remove the singularity found for θ_i and Φ_i at $x_i = 0$. These variables are defined as,

$$l_i = \frac{z}{2}(r_i - q_i) - (r_i - 1) \tag{3-35}$$

$$\theta_i' = \frac{\theta_i}{x_i} \tag{3-36}$$

$$\Phi_i' = \frac{\Phi_\text{i}}{x_i} \tag{3-37}$$

3.2.3.1. *Pros & Cons*

- ☑ Like Wilson and NRTL there is a built in temperature dependence that is assumed valid over moderate ranges.
- ☑ Like the NRTL equation UNIQUAC can represent LLE; however it only requires two adjustable parameters per binary (versus the three needed for NRTL).
- ☑ Given its theoretical basis, the UNIQUAC model is generally accepted as being more widely applicable (there is a greater *comfort-factor* in its use), and it is capable of representing mixtures exhibiting different types of interactions between molecules (possibly superior representation of mixtures of widely different molecular sizes).
  - ☒ However the UNIQUAC equation is more complicated, and has been shown to be less accurate than the NRTL and Wilson equations (Malanowski, et al., 1992). Likely much of its support stems from its use in group contribution methods such as UNIFAC, where there is a great deal of effort being expended.

3.2.4. General Discussion

Wilson's concept of local composition has resulted in several successful methods that are still widely used in industry today (i.e. the Wilson, NRTL, and UNIQUAC equations). Given the importance of these methods, much effort has been spent on trying to understand Wilson's use of "local composition" and the equations based on it (McDermott, et al., 1977; Maurer, et al., 1978; Mollerup, 1981; Kemény, et al., 1981; Hu, et al., 1983; Hoheisel, et al., 1984; Mansoori, 1985; Góral, 1995). Given UNIQUAC's stronger theoretical foundation it is often viewed in a more favourable light than the Wilson or NRTL equations, although the method typically underperforms in comparison tests (Lafyatis, et al., 1989; Malanowski, et al., 1992).

Practitioners will often have their favourites for one reason or another; however choices should be made based on which method provides the best results on a case by case basis. When considering modelling applications on the other hand, generalizations are often tactfully employed since individual unit operations must use a single g^E-model to represent the phase behaviour of multicomponent streams in a consistent manner. A judicious choice must often be made; in the context of the pros and cons already stated for each method, these additional *guidelines* seem appropriate for practical use:

> If it is known a priori, that LLE does not exist, then either the Wilson or NRTL equation can be used with equivalent results. Wilson describes steep ascending slopes towards γ_i^∞ better then NRTL; else, preference is often given to the NRTL equation in view of its greater flexibility in representing the intricacies of various g^E shapes.
>
> If it is known, or not known, a priori, that LLE exists then either NRTL or UNIQUAC can be employed. Preference in this case is usually given to NRTL, given that its increased flexibility (Morisue, et al., 1972; Walas, 1985) tends to give superior results over those of UNIQUAC (Lafyatis, et al., 1989; Malanowski, et al., 1992).
>
> - Although the Wilson equation cannot predict phase splitting, it can still predict realistic VLE behaviour inside the miscibility gap without stability tests (Rarey, 2009-2010).

3.3.Group Contribution Methods

The idea that solutions can be represented by a mixture of molecular-groups is generally credited to Langmuir (Deal, et al., 1968; Fredenslund, et al., 1975; Kojima, et al., 1979), with contributing work performed by Brønsted & Koefoed (Hála, 1978). This concept was later formalized into the *solution of groups* (SOG) concept by Wilson & Deer (1962), and provides a common foundation on which all group contribution g^E-models are based. Given the industrial importance of these methods, a derivation loosely following Hála (1978) will be presented here.

A solution can conceptually be considered a solution of pure components $(X, Y, \cdots Comp)$ at constant temperature and pressure:

$$n_1\mathbb{C}_1 + n_2\mathbb{C}_1 + \cdots + n_\# Comp; \quad [T, P] \tag{3-38}$$

On the other hand, each of the components can be conceptually *fragmented* and represented as a combination of molecular-groups $(\mathbb{G}_1, \mathbb{G}_2, \cdots Group)$:

$$\begin{aligned} \mathbb{C}_1 &= \nu_1^{(1)}\mathbb{G}_1 + \nu_2^{(1)}\mathbb{G}_2 + \cdots + \nu_*^{(1)} Group \\ \mathbb{C}_2 &= \nu_1^{(2)}\mathbb{G}_1 + \nu_1^{(2)}\mathbb{G}_2 + \cdots + \nu_*^{(2)} Group \\ &\vdots \\ Comp &= \nu_1^{(\#)}\mathbb{G}_* + \nu_2^{(\#)}\mathbb{G}_* + \cdots + \nu_*^{(\#)} Group \end{aligned} \tag{3-39}$$

where $\nu_{Group}^{(Comp)}$ represents the number of molecular group types in each pure component. Therefore an overall balance can be written in terms of these molecular groups:

$$\underbrace{[n_1\mathbb{C}_1 + n_2\mathbb{C}_1 + \cdots + n_{\#}Comp]}_{Soluiton\ of\ pure\ components} = \underbrace{[\nu_1\mathbb{G}_1 + \nu_2\mathbb{G}_2 + \cdots + \nu_k\mathbb{G}_\mathrm{k} + \cdots + \nu_{*}Group]}_{Soluiton\ of\ pure\ molecular\ groups} \tag{3-40}$$

where ν_k represents the total number of moles of the k^{th} group in the mixture:

$$\nu_k = n_1\nu_k^{(1)} + n_2\nu_k^{(2)} + \cdots + n_{\#}\nu_k^{(\#)} \tag{3-41}$$

According to Equation (2-29), Equation (2-49), and Equation (2-52) the liquid phase activity for such a process can then be calculated as

$$\ln\gamma_i = \sum_{k=1}^{\#Groups} \nu_k^{(i)}\left[\ln\Gamma_k^{(Mix)} - \ln\Gamma_k^{(i)}\right] \tag{3-42}$$

where

$\nu_k^{(i)}$: is the number of groups of the k^{th} type in the i^{th} molecule
$\ln\Gamma_k^{(Mix)}$: is the group activity of the k^{th} type in the mixture (SOG)
$\ln\Gamma_k^{(i)}$: is the group activity of the k^{th} type in a pure liquid of the i^{th} molecule

An illustration of Equation (3-42) is given for the system ethanol (1) + acetaldehyde (2) (see Figure 3), where ethanol consists of $1(\mathrm{CH_3}) + 1(\mathrm{CH_2}) + 1(\mathrm{OH})$ and acetaldehyde consists of $1(\mathrm{CH_3}) + 1(\mathrm{CHO})$. Therefore, in the mixture, a $(\mathrm{CH_3})$ group will encounter $(\mathrm{CH_3})$, $(\mathrm{CH_2})$, (OH), and the groups of acetaldehyde, meaning that the partial molar excess Gibbs energy of ethanol in the mixture then depends on the number of each group-type available in the mixture—as shown in Equation (3-38) through Equation (3-41), this mixture can be approximated as a solution of groups.

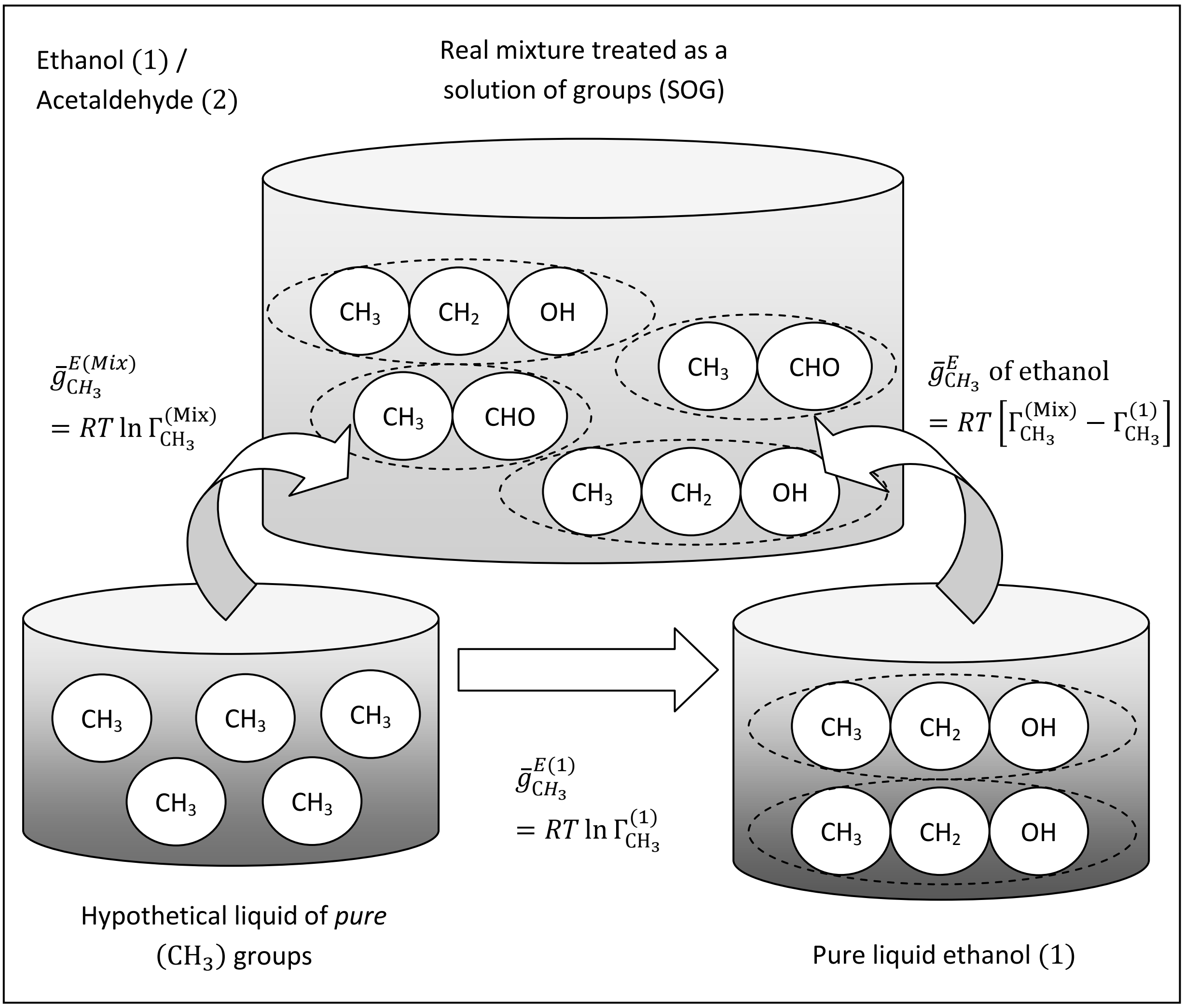


Figure 3 Illustration relating the partial molar excess Gibbs energy of (CH_3) groups in pure liquid ethanol (1), a real solution of groups where acetaldehyde is component (2), and a hypothetical liquid of pure (CH_3) groups. The partial molar excess Gibbs energy for (CH_3) in ethanol is then calculated by taking the difference between the other two paths.

The partial molar excess Gibbs energy of ($\mathrm{CH_3}$) groups, therefore, can be calculated relative to a hypothetical solution of pure ($\mathrm{CH_3}$) groups. Where $\bar{g}^{E(CH_3)}_{CH_3} = RT \ln \Gamma^{(CH_3)}_{CH_3} = 0$ is the partial molar excess Gibbs energy of the ($\mathrm{CH_3}$) groups in a hypothetical liquid of pure ($\mathrm{CH_3}$) groups, and $\bar{g}^{E(Mix)}_{CH_3} = RT \ln \Gamma^{(Mix)}_{CH_3}$ is the partial molar excess Gibbs energy of the ($\mathrm{CH_3}$) groups in the mixture (i.e. the solution of groups). Where the partial molar excess Gibbs energy of the ($\mathrm{CH_3}$) groups in pure ethanol, given by $\bar{g}^{E(1)}_{CH_3}$, will be different from $\bar{g}^{E(CH_3)}_{CH_3}$, because even in pure ethanol the ($\mathrm{CH_3}$) groups will still encounter a mixture of ($\mathrm{CH_3}$), ($\mathrm{CH_2}$), and (OH) groups in the quantities they appear in pure ethanol. The activity of the ($\mathrm{CH_3}$) groups in pure ethanol, $\ln \Gamma^{(1)}_{\mathrm{CH_3}} = \mathrm{g}^{\mathrm{E(1)}}_{\mathrm{CH_3}}/\mathrm{RT}$, is therefore not zero, where the superscript (1) indicates pure component (1).

The difference that is desired, however, is the effect of mixing the ($\mathrm{CH_3}$) groups in ethanol with component (2) (acetaldehyde in this case), relative to pure ethanol. This can be obtained by simply subtracting $\bar{g}^E$ of the ($\mathrm{CH_3}$) groups in a liquid of pure ethanol from the $\bar{g}^E$ of the ($\mathrm{CH_3}$) groups in the mixture: $\bar{g}^{E(Mix)}_{CH_3} - \bar{g}^{E(1)}_{CH_3}$. It is therefore possible to determine the excess Gibbs energy of ethanol in the mixture by summing up the partial molar contributions (via Equation (3-42)).

3.3.1. ASOG Method

Wilson & Dear (1962) showed that the SOG construct could successfully be used for making estimates from limited amounts of data – a large monetary incentive for industry. This early success paved the way for the analytical solution of groups' equation (ASOG), arguably the first method sufficiently developed for direct use by industry (Kojima, et al., 1979).

Like Wilson & Dear it was assumed that the excess chemical potential $(RT \ln \gamma_i)$ is comprised of two separate contributions:

$$\ln \gamma_i = \underbrace{\ln \gamma_i^C}_{\substack{Combinatorial \\ contribution}} + \underbrace{\ln \gamma_i^R}_{\substack{Residual \\ contribution}} \tag{3-43}$$

where like UNIFAC, the *combinatorial* term accounts for differences in molecular size, and the *residual* term accounts for the intermolecular forces of the molecules. Differences in molecular size were taken into account using a modified Flory-Huggins relation:

$$\ln \gamma_i^C \equiv \ln \gamma_i^{FH} = \ln \mathrm{R}_\mathrm{i}^* + 1 - \mathrm{R}_\mathrm{i}^* \tag{3-44}$$

where the convenience variable R_i^* is defined as (Naidoo, 2007),

$$\mathrm{R}_\mathrm{i}^* \equiv \frac{\sum_{k=1}^{Group} \nu_{ki}}{\sum_{k=1}^{Group} \sum_{j=1}^{Comp} \nu_{kj} x_j} \quad [=] \text{ Pure-component group ratio of the SOG} \tag{3-45}$$

and the residual term is calculated using the fundamental SOG relationship that was derived in the preceding chapter (see Equation (3-42)):

$$\ln \gamma_i^R = \sum_{k=1}^{Group} \nu_{ki} \left(\ln \Gamma_\mathrm{k}^{(Mix)} - \ln \Gamma_\mathrm{k}^{(i)} \right) \tag{3-46}$$

The Wilson equation was adapted to the SOG construct and used to calculate the pure-group activities and the activities of the groups in solution:

$$\Gamma_\mathrm{k}^{(Mix)} = 1 - \ln\left(\sum_{\mathrm{m}=1}^{\mathrm{Groups}} \mathrm{X_m} \Psi_\mathrm{mk} \right) - \sum_{\mathrm{n}=1}^{\mathrm{Groups}} \left(\frac{\mathrm{X_n} \Psi_\mathrm{kn}}{\sum_{\mathrm{m}=1}^{\mathrm{Groups}} \mathrm{X_m} \Psi_\mathrm{mn}} \right) \tag{3-47}$$

$$\ln \Gamma_\mathrm{k}^{(i)} = 1 - \ln\left(\sum_{\mathrm{m}=1}^{\mathrm{Groups}} \mathrm{X_{mi}^{(i)}} \Psi_\mathrm{mk} \right) - \sum_{\mathrm{n}=1}^{\mathrm{Groups}} \left(\frac{\mathrm{X_{ni}^{(i)}} \Psi_\mathrm{kn}}{\sum_{\mathrm{m}=1}^{\mathrm{Groups}} \mathrm{X_{mi}^{(i)}} \Psi_\mathrm{mn}} \right) \tag{3-48}$$

where

$$\mathrm{X_m} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{Comp}} \mathrm{x_i \nu_{mi}}}{\sum_{\mathrm{i=1}}^{\mathrm{Comp}} \sum_{\mathrm{m=1}}^{\mathrm{Group}} \mathrm{x_i \nu_{mi}}} \;\; [=] \;\; \text{Mixture group-fractions} \tag{3-49}$$

$$\mathrm{X_{mi}^{(i)}} = \frac{\nu_{ki}}{\sum_{\mathrm{m=1}}^{\mathrm{Groups}} \nu_{mi}} \;\; [=] \;\; \text{Pure-group fractions} \tag{3-50}$$

and the group interaction energies are obtained from

$$\Psi_{\mathrm{mk}} = \exp\left(\zeta_{mk} + \frac{\xi_{mk}}{T}\right) \tag{3-51}$$

where Ψ_{mk} are the group interaction parameters attributed to groups m and k, where $\Psi_{mk} \neq \Psi_{km}$ and $\Psi_{mm} = 1$. So here the temperature dependence is explicitly correlated using the two group interaction parameters ζ_{mk} and ξ_{mk} (Tochigi, et al., 1976; Tochigi, et al., 1977; Kojima, et al., 1979) whereas LC-equations *typically* only make use of an implied temperature relationship.

3.3.2. UNIFAC Method

The UNIFAC method is similar to the ASOG method except for the equations used to represent the combinatorial and residual contributions of the solute-activity. The UNIQUAC method was adopted as an alternative to the Flory-Huggins and Wilson adaptations used in ASOG; it was from this extension of UNIQUAC that gave UNIFAC its name (UNIQUAC Functional-group Activity Coefficient model). Given the promising results of UNIQUAC at the time, the decision was made to see if the benefits would extend into the SOG concept (Fredenslund, et al., 1975).

The group fractions in the mixture (using Equation (3-49)) and the pure component group fractions (using Equation (3-50)) are calculated as in ASOG, and likewise with the residual term, but without the built-in temperature dependence:

$$\Psi_{\mathrm{mk}} = \exp\left(-\frac{(\varepsilon_{mk} - \varepsilon_{kk})}{T}\right) \tag{3-52}$$

As in the ASOG method, the application of UNIQUAC to SOG requires adaptations to the standard UNIQUAC terms. The pure component properties are transformed into the sum of the group volume and area parameters (R_k and Q_k):

$$r_i = \sum_{\mathrm{k=1}}^{\mathrm{Group}} \mathrm{\nu_k^{(i)} R_k} \;\; [=] \;\; \text{pure group volumes} \tag{3-53}$$

$$q_i = \sum_{\mathrm{k=1}}^{\mathrm{Group}} \mathrm{\nu_k^{(i)} Q_k} \;\; [=] \;\; \text{pure group surface areas} \tag{3-54}$$

and the group activity coefficient $\Gamma_{\mathrm{k}}^{(Mix)}$ in the mixture, and pure group activity coefficients $\Gamma_{\mathrm{k}}^{(i)}$ are represented as

$$\ln \Gamma_k^{(Mix)} = Q_k \left[1 - \ln\left(\sum_{m=1}^{Group} \Theta_\mathrm{m} \Psi_\mathrm{mk} \right) - \sum_{m=1}^{Group} \frac{\Theta_\mathrm{m} \Psi_\mathrm{km}}{\sum_{n=1}^{Group} \Theta_\mathrm{n} \Psi_\mathrm{nm}} \right] \tag{3-55}$$

$$\ln \Gamma_k^{(i)} = Q_k \left[1 - \ln\left(\sum_{m=1}^{Group} \Theta_\mathrm{m}^{(i)} \Psi_\mathrm{mk} \right) - \sum_{m=1}^{Group} \frac{\Theta_\mathrm{m}^{(i)} \Psi_\mathrm{km}}{\sum_{n=1}^{Group} \Theta_\mathrm{n}^{(i)} \Psi_\mathrm{nm}} \right] \tag{3-56}$$

where

$$\Theta_\mathrm{m} = \frac{Q_\mathrm{m} X_\mathrm{m}}{\sum_{n=1}^{Group} Q_\mathrm{n} X_\mathrm{n}} \tag{3-57}$$

$$\Theta_\mathrm{m}^{(i)} = \frac{Q_\mathrm{m} X_\mathrm{mi}^{(i)}}{\sum_{n=1}^{Group} Q_\mathrm{n} X_\mathrm{mi}^{(i)}} \tag{3-58}$$

3.3.3. General Discussion

Both ASOG and UNIFAC have been amply tested in literature, but considerably more effort has been expended on the UNIFAC model (and its modifications). In this Chapter the general differences between the two methods will be discussed: concentration dependence, temperature dependence, and group definitions.

3.3.3.1. *Concentration Dependence*

Since the formalization of the SOG concept in 1962, improvements have been made by incorporating better concentration dependent relationships. Up until the introduction of the Wilson equation in 1964, industry had to rely on cumbersome algebraic and empirical equations (usually derived on a case-by-case basis); given that the Wilson equation was proving superior to other methods in practice at the time, it seems quite logical that the Wilson equation was extended to the SOG concept. This extension (ASOG) was being formulated as early as 1968 (Deal, et al., 1968), and eventually was further developed with practitioners in mind (Kojima, et al., 1979).

As previously discussed, the success of Wilson paved the way for other local-composition models like NRTL (Renon, et al., 1968) and UNIQUAC (Abrams, et al., 1975). In view of the benefits of UNIQUAC (see Chapter 3.2.3), and given that the expression is written as a combination of combinatorial and residual contributions, the expression naturally lends itself to be extended using the SOG concept (Fredenslund, et al., 1975). Where in-kind to the ASOG method, this additional adaptation (UNIFAC) was further developed with practitioners in mind (Fredenslund, et al., 1977).

These *original* methods, as initially presented in literature, seem to give about equal reliability. The UNIFAC method, however, was singled out for further development (Gmehling, 2009). One explanation is that this decision was based solely on the broader applicability of the UNIQUAC equation as compared to the Wilson model at the time (e.g. the UNIQUAC model was more robust, and was supported by a sounder theoretical basis rooted in statistical mechanics). As for the Wilson model's inability to represent LLE, it becomes a "moot-point" when the SOG construct is applied to the model (Rizzi, et al., 1981).

3.3.3.2. *Combinatorial Contribution*

Early work to address the predictive results for asymmetric systems led to the following UNIFAC modification of the combinatorial expression (Weidlich, et al., 1987):

$$\ln \gamma_i^C = 1 - \Phi_i' + \ln \Phi_i' - 5q_i\left(1 - \frac{\Phi_i}{\theta_i} + \ln\frac{\Phi_i}{\theta_i}\right) \tag{3-59}$$

where the ¾ term was introduced into Φ_i' to correct for the overly large combinatorial contribution of the original UNIFAC expression (Thomas, et al., 1984):

$$\Phi_i' = \frac{r_i^{3/4} x_i}{\sum_{j=1}^{Comp} r_j^{3/4} x_j} \tag{3-60}$$

3.3.3.3. *Temperature Dependence*

A major difference between the ASOG method and the original UNIFAC equation was their treatment of the group-temperature-dependence. The temperature dependence of the group contribution parameters of the ASOG model were well evaluated by Derr and Deal (1969), Ronc and Ratcliff (1971), and Kojima *et al* (Tochigi, et al., 1976; Tochigi, et al., 1977; Kojima, et al., 1979; Tochigi, et al., 1981). Changes were later made, however, to the UNIFAC method to include temperature-dependent interaction parameters (amongst other modifications) by Weidlich and Gmehling (1987):

$$\Psi_{\mathrm{mk}} = \exp\left[-\left(\underbrace{(\varepsilon_{mk}^a - \varepsilon_{kk}^a)}_{a_{mk}} + \underbrace{(\varepsilon_{mk}^b - \varepsilon_{kk}^b)}_{b_{mk}}T + \underbrace{(\varepsilon_{mk}^c - \varepsilon_{kk}^c)}_{c_{mk}}T^2\right)\Big/T\right] \tag{3-61}$$

3.3.3.4. *Definition of Groups*

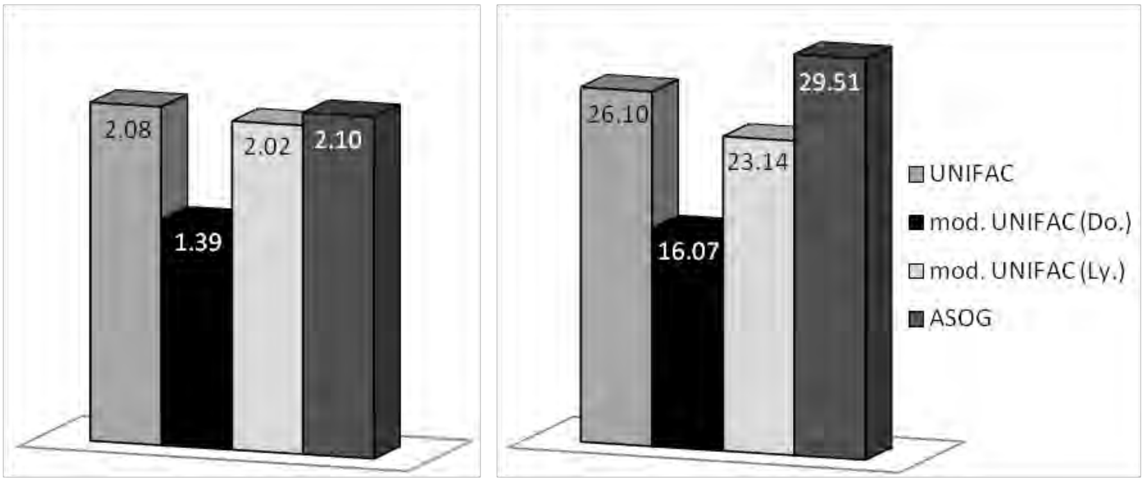
Another major difference between ASOG and UNIFAC relations are the group definitions used by each. This difference is especially boisterous considering that the whole principle behind SOG rests on the *additivity* of functional groups; given the historic ambiguity of defining functional groups, the specific details concerning the differences are inconsequential.

It is of interest to note, however, that the lack of theoretical basis has been given some attention (Wu, et al., 1991)*;*[a, b] *however, it is a generally accepted viewpoint that the choice of groups is largely arbitrary.*

3.3.3.5. *Present Status*

The *modified UNIFAC* (Dortmund) method (Weidlich, et al., 1987) encompasses the changes to the combinatorial and temperature-dependent expressions that were covered in the preceding Chapters (Equation (3-59) thru Equation (3-61)). This modification has essentially narrowed the fundamental differences between the ASOG and UNIFAC methodologies. The modified combinatorial expression of mod. UNIFAC (Do.) basically provides a middle-of-the-road representation between these two original methods, and the method was later extended to include group interaction parameters quadratic in temperature. The largest difference left between ASOG and modified UNIFAC (Do.) is then the quality of the data used to determine their parameters; clearly in this regards the UNIFAC and modified UNIFAC (Do.) methods have exceedingly benefited by continuous improvement (Weidlich, et al., 1987; Gmehling, et al., 1993; Gmehling, et al., 1993; Gmehling, et al., 1998; Lohmann, et al., 2001; Gmehling, et al., 2002; Fischer, et al., 2007; Gmehling, 2009).

Figure 4 and Figure 5 show typical results between calculated values using these group methods and experimental data. The first figure shows calculated absolute deviations and relative deviations for activity coefficients at infinite dilution, while the second figure shows calculated absolute deviations in vapour mole fraction y (%), temperature T (K), and pressure P (kPa). When limiting the discussion to only the group contribution methods covered in this Chapter, mod. UNIFAC (Do.) proves superior in all categories (see Figure 4 and Figure 5). For comparison, Figure 5 also includes results obtained by a direct fit of the VLE data using the 2-parmeter UNIQUAC equation. These calculated results then represent the experimental scatter of the data, and the inability of the 2-parameter model to correctly fit the experimental findings; nonetheless the UNIQUAC model, fitted to each experimental dataset individually, proves superior to the predicted results obtained using the group contribution methods.



absolute: $\left|\gamma_{i,pred}^{\infty} - \gamma_{i,exp}^{\infty}\right|$

relative: $\frac{\left|\gamma_{i,pred}^{\infty} - \gamma_{i,exp}^{\infty}\right|}{\gamma_{i,exp}^{\infty}} \cdot 100\%$

Figure 4 Absolute and relative deviations between experimental and predicted activity coefficients at infinite dilution for 13,500 data points (Lohmann, et al., 2001).

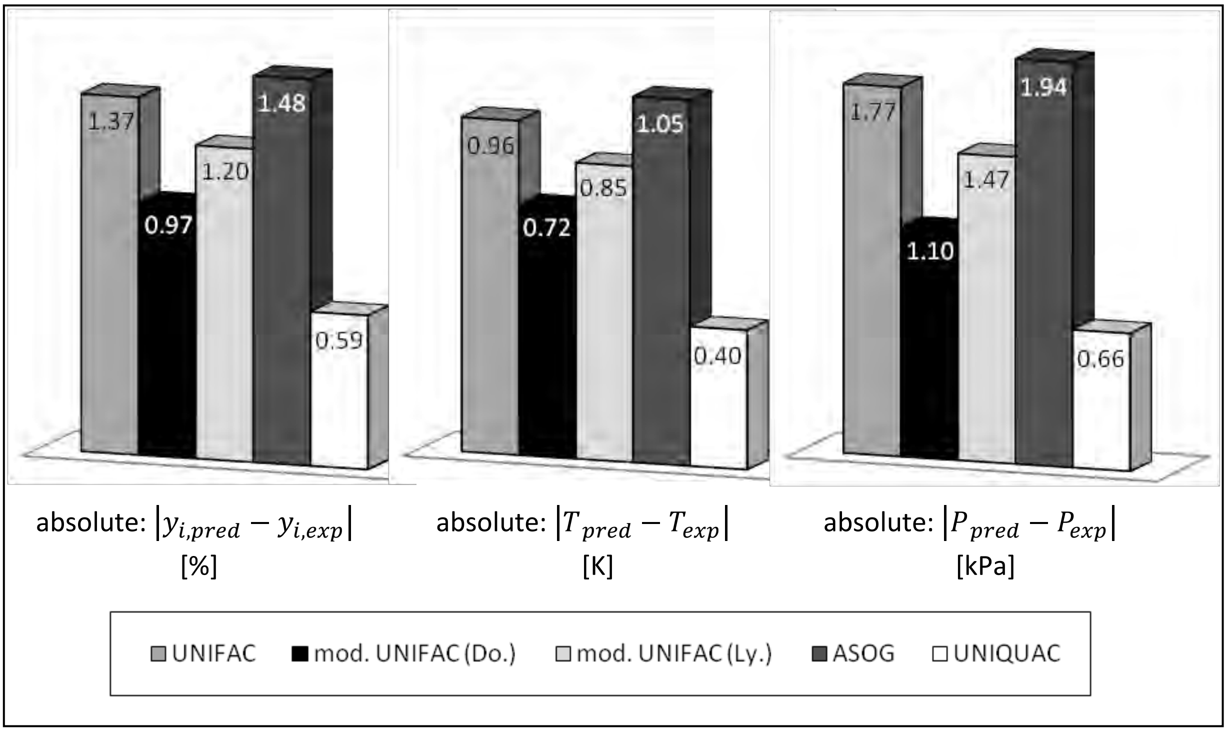


Figure 5 Calculated absolute deviations in vapour mole fraction, temperature T, and pressure P for 2,400 thermodynamically consistent isothermal / isobaric VLE datasets (Lohmann, et al., 2001).

4. Proposed Method

Simulations can be extremely useful tools used by industry, both for designing new plants (or processes) and for optimizing existing operations. The success of these simulations rests heavily on how accurate thermodynamic properties can be represented; their calculation and understanding are of fundamental importance. This is easier said than done, for there are a number of “show stoppers” that prevent most practitioners from using process simulators effectively. Determining which method to use, handling missing model parameters, unavailability of experimental data, and interpreting and spotting the significance of results are just a few of the obstacles one has to navigate, in order to obtain results which are sufficiently accurate and useful in practice.

To meet industrial needs, a new method is being proposed which should enhance the usability of group contribution methods such as UNIFAC. Where instead of regressing experimental data directly using LC-models such as the NRTL equation (i.e. component-based methods), predictions of the experimental data will be corrected for instead. This would eliminate the practice of fitting missing model parameters to predictive results when no experimental data of sufficient quality can be found; instead, predictive methods can be used directly in process simulations. To this end, the formulation of the proposed *hybrid method* necessitates the incorporation of the following features:

1. The method needs to retain the same predictive results as the base-GC model used in the hybrid model; so that when no experimental data exists, practitioners can use predictive results directly without the danger of converting the results to a component-based model (i.e. eliminate the potential for the “loss in translation”).
2. The hybrid method should also be kept as simple as possible to facilitate its incorporation into existing industrial-software; either by the software designers themselves, or by practitioners using the software to solve real-world problems.

A simple augmentation meeting the above criteria is the following:

$$\ln \gamma_i = \overbrace{\underbrace{\ln \gamma_i^C}_{\substack{Combinatorial\\contribution}} + \underbrace{\ln \gamma_i^R}_{\substack{Residual\\contribution}}}^{base-GC\ model} + (\text{LC-Model}) = \ln \gamma_i^{GC} + \ln \gamma_i^{LC} \tag{4-1}$$

The end result is the proposed hybrid model, which layers a component-based correction on top of the base-GC model used. This essentially creates a layer-cake method, which can conceivably be extended by layering additional corrections on top of one another (as long as no overlap exists between the various layers). Instead of regressing the experimental data directly, the component-based correction would regress the errors between the predictions and any available experimental data. This concept is illustrated in Figure 6 below using the mod. UNIFAC (Do.) method, for the system n-hexane (1) + 4-methylpyridine (2), where the hybrid model would adjust the base-prediction (dotted line in Figure 6) to match the experimental data points, by fitting the component-based correction to the error of the prediction.

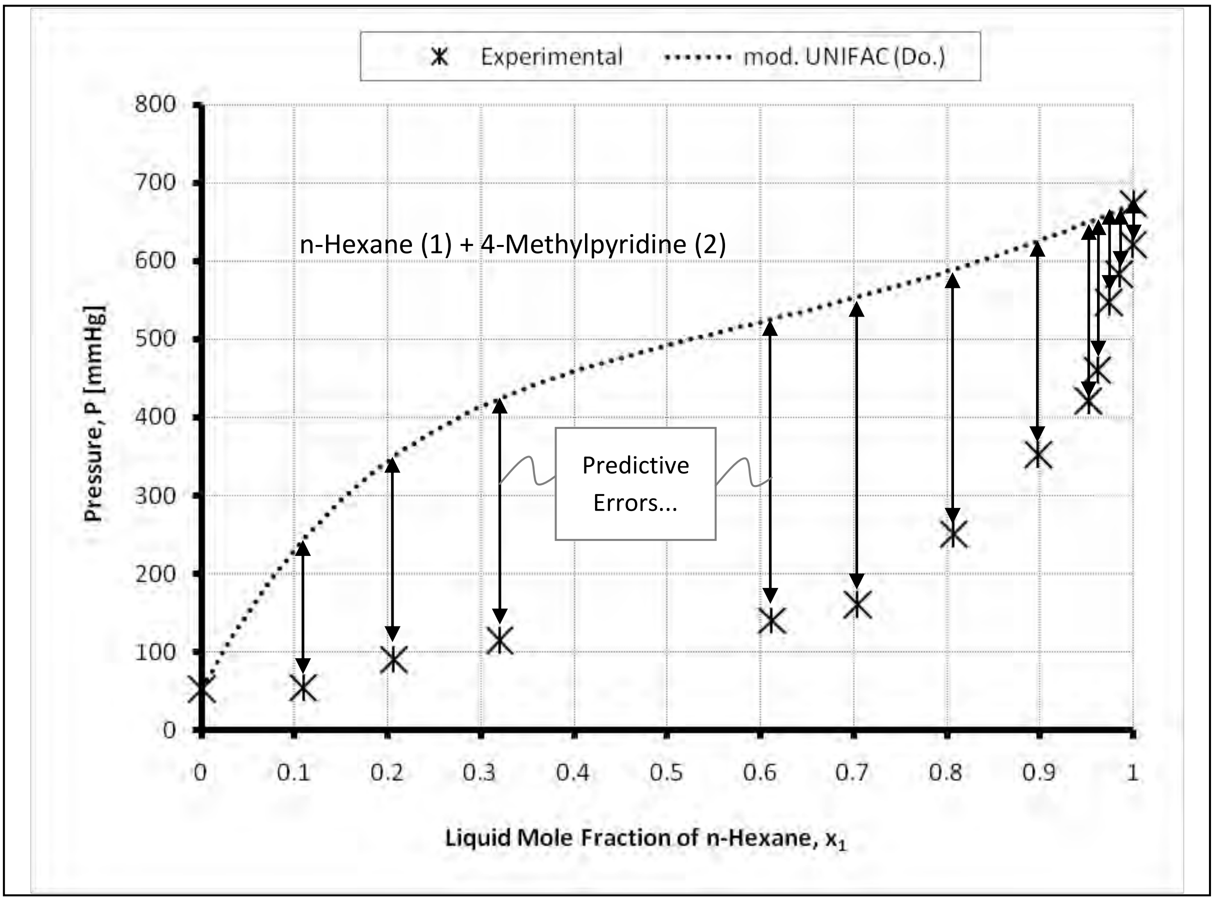


Figure 6 An illustration of the hybrid concept using DDBVLE dataset 13199 (n-hexane (1) + 4-methylpyridine (2) at 338.15 K), where the errors in the predictions using the base-GC method would be regressed using the model employed for the component-based correction, instead of regressing the experimental data directly (as is convention).

4.1. Combinatorial Term

Recent work performed by Moller (2010) compared various combinatorial expressions; the study evaluated the combinatorial expressions of Guggenheim-Staverman, mod. UNIFAC, and the free volume expression GK-FV. It was observed that GK-FV provided superior results for non-polymer solutes, which is surprising since the concept of free volume (FV) is almost always associated with polymer applications (Moller, 2010). At face value, this suggests that free volume effects should be included in the contributions of the combinatorial:

$$\ln \gamma_i^{C(Cav)} = \ln \gamma_i^{C(FV)} + \ln \gamma_i^{Cav} \tag{4-1}$$

$$\ln \gamma_i^{Cav} = \left(\frac{\Phi_\mathrm{i}^\mathrm{FV}}{x_i} - \frac{\Phi_\mathrm{i}^{\mathrm{FV}\star}}{x_i} \right) \tag{4-2}$$

where the free volume represents the volume not occupied by the molecules in solution, and are used to determine the free volume fractions,

$$\Phi_{\mathrm{i}}^{\mathrm{FV}} = \frac{x_i V_i^{FV}}{\sum_{\mathrm{j=1}}^{\mathrm{Comp}} x_j V_j^{FV}} \tag{4-2}$$

$$\Phi_{\mathrm{i}}^{\mathrm{FV}\star} = \frac{x_i V_i^{FV\star}}{\sum_{\mathrm{j=1}}^{\mathrm{Comp}} x_j V_j^{FV\star}} \tag{4-3}$$

$$V_i^{FV} = \upsilon_{\mathrm{i}} - \Phi_{\mathrm{i}} \tag{4-4}$$

$$V_i^{FV\star} = (V_i)^{2/3} - (V_i^{\star})^{2/3} \tag{4-5}$$

where V_i is the liquid molar volume and $V_i^{\star}$ is the van der Waals volume according to Bondi (1964).

This relationship (see Equation (4-1)) was further modified in order to correct for under predictions caused by the cavitation contribution when the size of the solute becomes larger than the size of the solvent (i.e. the case of large in small). To overcome this limitation, a modified GK-FV was proposed by Moller (2010). This included a multiplication factor to correct for the under prediction, and a step function to ensure that an activity coefficient of unity is obtained for mixtures of identical size and shape. Developed for binary systems, where component-1 is the solvent and component-2 is the solute:

$$\ln \gamma_2^{Cav}\Big|_{corr} = \ln \gamma_2^{Cav} * [1.15 - 0.15 * \exp(-0.5(q_1 - q_2)^2 - 0.5(r_1 - r_2)^2)] \tag{4-6}$$

The modification provides a far better prediction for large solutes in small saturated hydrocarbon solvents, while maintaining near-equivalent results for small solutes in larger saturated hydrocarbons (Moller, 2010). The combinatorial expression was therefore incorporated into the present study of the proposed hybrid method, so that the ability of the equation to describe non-hydrocarbon solvent-systems could be evaluated. From this point forward, Equation (4-1) through Equation (4-6) will be referred to as the MRR combinatorial.

4.2. Hybrid Tuning

Given that the UNIQUAC equation tends to greatly over predict two-phase regions (Lafyatis, et al., 1989) this study opted to use the NRTL equation to develop the concept of the hybrid group contribution model. Furthermore, the increased flexibility of the NRTL equation seemed amply suited for the tuning portion of the hybrid method. This choice was later expanded to include the Wilson equation, given that it is best suited to represent sharp inclines in the dilute regions of the phase envelope

5. Results and Discussion

5.1. Evaluation of the Solution of Groups Concept

Process simulators have become commonplace tools in industry; whether being used for design, debottlenecking, or troubleshooting chemical operations, their value relies on the sufficiently precise description of various thermophysical properties. Liquid-phase activity coefficients are one such property. They are particularly important for the description of many separation processes, where knowledge of component splits between various phases is required (distillation and separators are such examples). One of the most common ways of obtaining this description is via activity coefficients from component-based models such as Wilson, NRTL, or UNIQUAC.

To reach an adequate representation of the behaviour of real mixtures, binary interaction parameters of these component-based models are typically fitted to experimental data or to predictive results obtained from group contribution (GC) methods such as UNIFAC. These two approaches employ the following steps:

1. **Component Approach:** Experimental data → fitting binary interaction parameters between whole molecules for each *individual* binary mixture.
2. **Group Approach:** Experimental data → fitting of binary interaction parameters between molecular groups to *several* datasets (simultaneously) → the prediction of data for each binary mixture of components → fitting of binary interaction parameters to these predictions.

Although both approaches can be used in principle, the second case would have the advantage of being able to describe mixtures for some binary combinations where no experimental data are available; this is very convenient in cases where the mixture contains several similar components from one chemical family plus additional component types. Here multiple systems exhibit similar behaviour, and there is hardly any reason to measure data for every binary system. Instead, it is much more efficient to translate the available information into a subset of molecular groups capable of representing those mixtures for which data are not available; given that the "available" experimental information was of *sufficient quality*, these same molecular groups can then be used to characterize similar mixture behaviour not yet measured.

The solubility of n-alkanes in some common solvent is one such example, where the solubility of every component would not need to be measured. Instead, measurements of two or three n-alkanes in the solvent would be enough to fit the maingroup interactions necessary for representing the trend in the series (given the availability of quality experimental data).

An example of this is shown in Figure 7, where the predicted relative volatilities[5] of n-alkanes in ethanol are compared against available experimental data. In this case, if the calculation of mixture behaviour was limited to only those systems where experimental data existed, then only the

---

[5] $\alpha_{12}[=]\ \text{relative volatility} = \frac{y_1/x_1}{y_2/x_2} = \frac{K_1}{K_2}$

behaviour of mixtures containing heptane, octane, or nonane in ethanol could be determined. The group contribution methodology, however, provides an approach where the n-alkanes (belonging to the same chemical family) and ethanol can be represented by a common subset of molecular groups. This makes it unnecessary to have measurements for every binary combination of n-alkanes in ethanol as a requisite; instead group contribution methods such as mod. UNIFAC (Do.) can be used to provide a sufficiently precise description of mixture behaviour for the missing binary combinations of n-alkanes in ethanol.

This is just one example illustrating how models based on the group contribution approach can represent a wide range of binary mixture behaviour with a limited number of parameters (i.e. it provides a means to expand existing-knowledge about a few systems to many other mixtures).

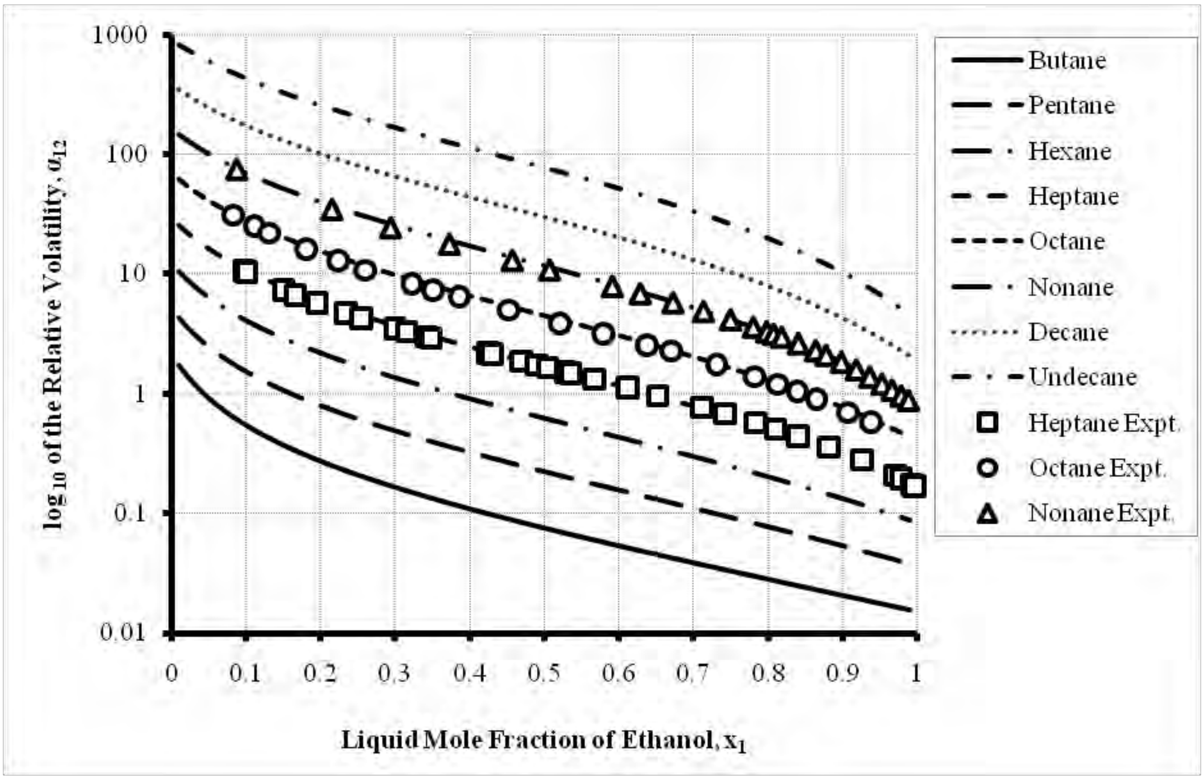


Figure 7 Relative volatilities (α_{12}) of different chain length n-alkanes (2) in ethanol (1), lines calculated using mod. UNIFAC (Do.) with Consortium 2008 parameters (The UNIFAC Consortium, 2008) against data taken from DDB (DDBST Software and Separation Technology GmbH, 2009).

The use of predictive models like UNIFAC during actual process simulation is mostly restricted to cases where a very large number of components are present in the mixture; in cases like this, component-based models would require an excessively large number of parameters (one for each individual binary combination of components in the mixture). However, there often exists a need to influence the description of individual binary mixtures without affecting the behaviour of several other binaries; in this case component approaches are very attractive, since the interaction between components "can be fitted" for each component-binary combination separately—the calculated behaviour of individual binary mixtures is treated *independently*.

For methods such as UNIFAC, the refitting of binary interactions between groups to match specific multicomponent mixtures is a much more difficult task than the regression of mixture data using component- models such as Wilson. The refits would be able to represent additional binary systems beyond those used for fitting purposes (compared to using component-models to regress the same data) and should result in better descriptions than using the official published parameters of the group contribution model. The modified interactions, however, would likely not be as reliable outside the range of mixtures they were fitted for.

Modifying the binary parameters between groups to obtain *specific results* for certain component-binaries, would affect any other mixture that contains the respective group combination/interaction—the varying of any one group parameter potentially influences the calculated behaviour of many other component-binaries at the same time. The fundamental need for practitioners to *easily* influence the description of individual binary systems, without inadvertently affecting others, has largely resulted in the practice of using and viewing group contribution methods as additional “data sources” (e.g. where no data exists) for the fitting of component-based models and not as models within the simulation itself.

The desire to match measured mixture data as accurate as possible encourages practitioners to fit specific component-binaries to experimental data using component-methods such as UNIQUAC. The main gain from this approach can be seen in Figure 8, where UNIQUAC parameters were independently fitted to individual datasets and compared against group contribution predictions based on the regression results of many thousands of datasets. To simplify discussion only absolute deviations in vapour fractions will be discussed, since parallel explanations can be made for the deviations in temperature and pressure that are shown.

The direct fitting of parameters for each binary system using UNIQUAC yields the lowest deviation; the 0.58% can be attributed to the scatter of experimental data and the inability of the two-parameter model to fit the datasets precisely. The slight increase to 0.88% obtained from modified UNIFAC (Do.) includes the short-comings of the group contribution method. It should be noted, however, that data from different authors for the same mixture often do not match.

The UNIQUAC regressions were fitted to each dataset individually, while mod. UNIFAC (Do.), as well as all other group contribution methods, must be regressed to all datasets at the same time. Discrepancies between the results of different authors, in this case, enter into the 0.88% of mod. UNIFAC (Do.), likewise for UNIFAC. Since the short-comings of the group contribution methods are *coupled* to the differences between authors, the increase to 1.41% obtained for UNIFAC represents the differences between the two group contribution methods: combinatorial modification, modified group fragmentations, types of data used in regression, and the addition of temperature dependent parameters in mod. UNIFAC (Do.).

While the deviations in the case of UNIFAC or mod. UNIFAC (Do.) are only slightly higher than in the case of directly fitting the datasets, it is nevertheless the current practice to stay within the component-based approach in order to be able to independently influence component-binary descriptions. This leads to several problems, Namely that the results of group contribution models may not be regressed with sufficient quality by the component-based models, whereby a “loss in translation” may occur when moving the results from group-based predictions to component-based models.

Group contribution models may give more realistic results as the different interactions are more localized and not “smeared” over the whole molecule.

In the subchapter below, these effects will be analyzed in detail.

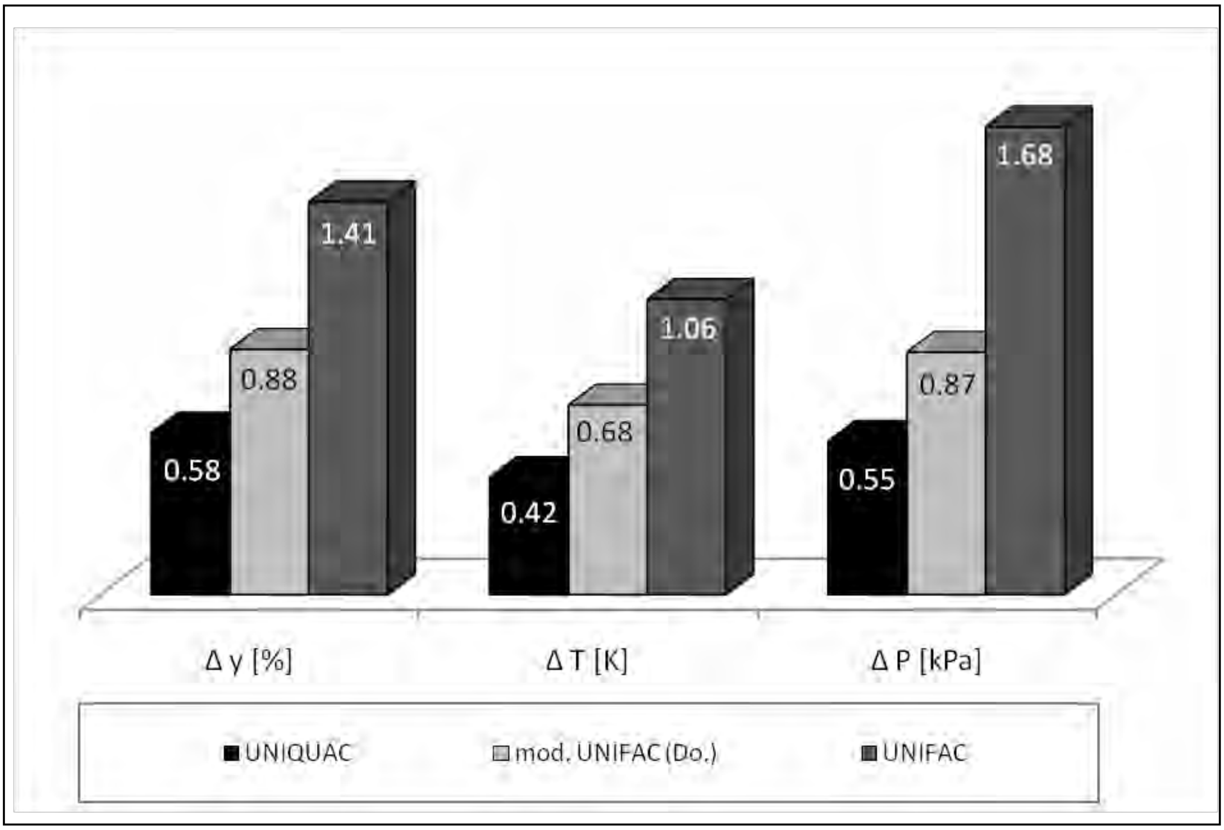


Figure 8 Relative or absolute deviations in vapour phase composition (y), equilibrium temperature in case of isobaric datasets (T), and equilibrium pressure for isothermal datasets (P) between experimental and predicted VLE data (2200 consistent data sets)— (Gmehling, 2009)

5.1.1. Ability to Represent the Real Behaviour of Binary Liquid Mixtures

Typically, only pure component and binary parameters are used in models for the description of the real liquid mixture behaviour. In large part, this is due to the scarcity and difficulty of measuring good ternary and higher data required for the fitting of any higher-order interactions between the molecules of multicomponent mixtures. These models, therefore, only differ in how they treat these binary interactions.

As mentioned in the preceding chapter, the need to influence the behaviour of specific component-binaries in multicomponent mixtures and to obtain the best agreement with “available” experimental data has led to the preferential use of component-based models. This preference results in the indirect use of group contribution methods as “data sources,” and can lead to a loss in translation when the predictions from group contribution models are regressed using component-based models. Therefore, one of the purposes of this work is to evaluate the possible combination of both approaches, making it necessary to determine which of the methodologies is best suited for representing the behaviour of real liquid mixtures—as a solution of whole molecules, or as a solution of molecular groups. The two approaches/methodologies will now be looked at in turn.

Component based models determine the total excess Gibbs energy of a mixture based on the interactions between individual molecules, whereas group contribution methods treat interactions between molecular groups that make up the molecules. The number of parameters in both situations is often the same, but the group approach may have a more physically sound basis.

Both approaches are depicted in Figure 9 using a mixture of n-hexane and ethanol for illustration purposes. The component-based approach is shown in Figure 9 (a), and makes no physical distinction of which molecular sites are involved in which types of intermolecular forces. The strong hydrogen "bonds" formed between the hydroxyl groups (—OH) of the ethanol molecules are therefore *lumped together* with the weak-dispersive forces occurring between the alkyl groups (—CH2, —CH3) of the n-hexane and ethanol molecules. Specific properties are therefore averaged over the *entire* molecular surface for each component-based interaction.

The group contribution approach on the other hand represents these interactions as occurring between molecular groups that make up the molecules (as shown in Figure 9 (b)). Although the treatment of interactions between whole molecules is *basically* correct (component-based approach), the strong H-bonding actually only exists between the hydroxyl groups of the ethanol molecules. By fragmenting molecules into a subset of molecular groups a much more physically realistic representation is obtained, where binary interactions between molecular groups (and not whole molecules) can be localized to the sites involved in specific interactions (as shown in Figure 9 (c)).

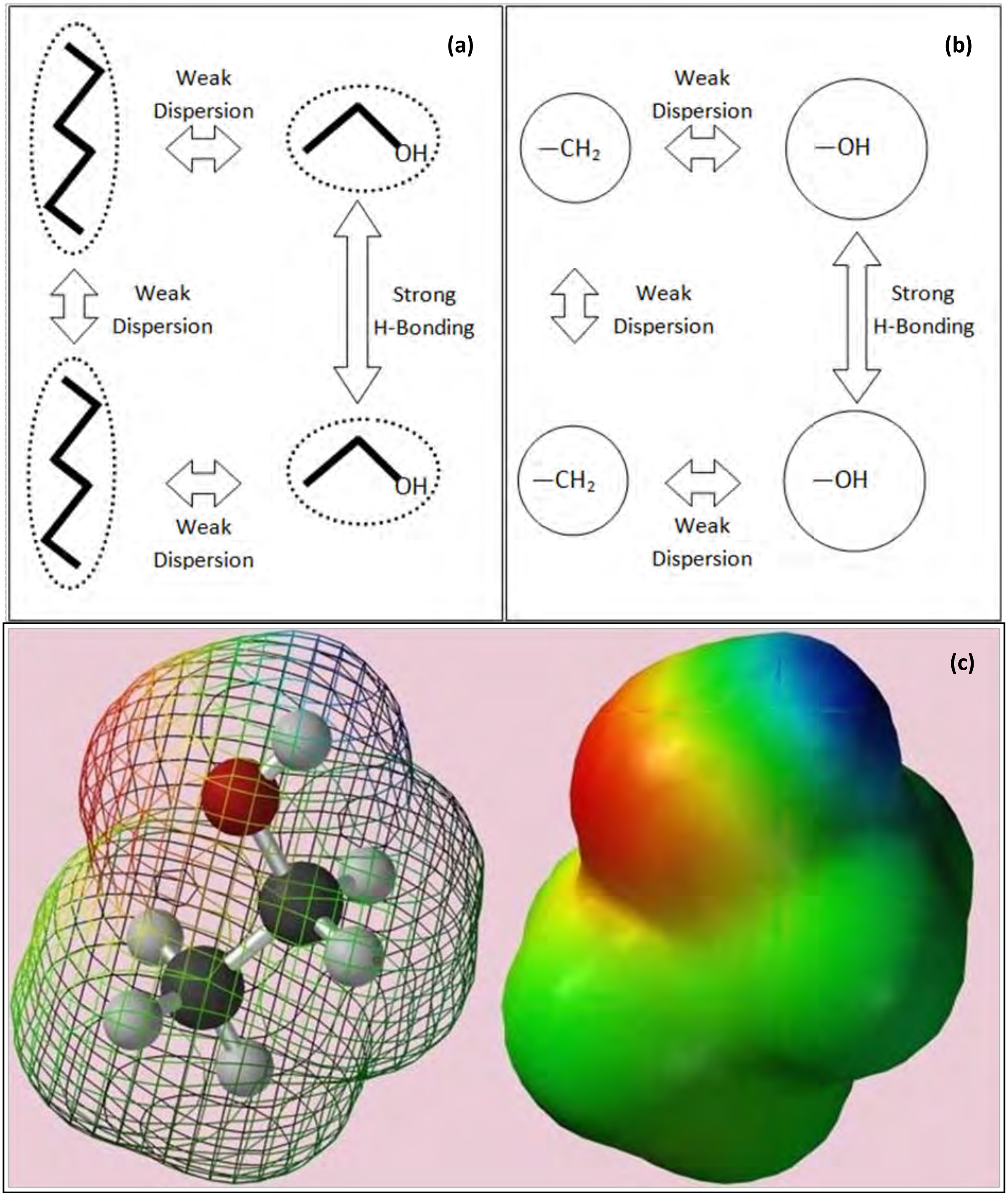


Figure 9 Component-based models treat interactions between whole molecules (a), while group contribution models use interactions between molecular groups for characterizing mixture behaviour (b). The group contribution approach therefore more closely resembles the detailed molecular models employed in modern molecular dynamics (MD) and Monte Carlo (MC) simulation, where the electrostatic potential of molecules are typically determined (e.g. (c) is of the ethanol molecule; red indicating a partial negative charged region and blue a partial positive charged region (Gallagher, 2007)).

5.1.2. Direct Fitting of Binary Interaction Parameters between Maingroups

In order to gain a better understanding of how well the group contribution approach represents the real behaviour of mixtures, binary interaction parameters between maingroups were fitted using the original UNIFAC method for each dataset separately and results were then compared to the regression results obtained from the component-based regressions using the Wilson, NRTL, and UNIQUAC models.

Binary isothermal data—Px(T)—were used for the evaluation, where the entire composition range for each binary dataset $(x_1\!: 0, 1)$[6] was divided into bins of 0.1 mole fractions each, and only those datasets having a minimum of 10 data points filling at least 5 of these bins were considered. The number of datasets were further restricted to those in which the authors' recorded the pure component vapour pressures of each mixture component (the end-points at infinite dilution). These restrictions removed many datasets of insufficient quality for the evaluation of the two approaches:

Each binary dataset covers the entire composition range so that the regression results represent how well each approach can characterize the curvature of each individual dataset.

Enough data points exist for each dataset, ensuring that adequate knowledge of the systems' pressure with changing composition exists for the fitting of interaction parameters with both approaches (to each individual dataset).

Using the authors' pure component vapour pressures avoids inconsistencies between data bank vapour pressure equation parameters and author measurements, where the calculated pure component vapour pressures can be adjusted to match those reported by the authors' (further elucidates the ability of each method to fit the curvature of the data).

It should be noted that the fitting of binary interaction parameters between groups requires that the group activity coefficient $\Gamma_{\mathrm{k}}^{(\mathrm{Mix})}$ in the mixture be different from the group activity coefficient $\Gamma_{\mathrm{k}}^{(\mathrm{i})}$ in the pure components of the mixture (recall Equation (3-47) and Equation (3-48)). If this was not the case then the residual contribution computed using Equation (3-46) would always result in a value of zero, and no binary interactions between groups could be determined. This situation may arise, for instance, in cases where both components are made up of identical groups with identical group frequencies (e.g. isomers). In these cases, the group contribution model will always predict ideal mixture behaviour independent of the group parameter values.

Using the above criteria the parameters of the Wilson, NRTL, UNIQUAC, and UNIFAC models were then fitted to each dataset individually, using the Simplex Nelder-Mead method (1965) to minimize the RMSD in pressure (the objective function, O.F., or referred to in this report as Pdev^2) for each dataset:

$$\mathrm{O.F.}\,(\mathrm{par1}, \mathrm{par2}, \dots) = \min \frac{1}{n} \sum_{i=1}^{n} \left(\frac{P_{exp,i} - P_{calc,i}}{P_{exp,i}} \right)^2 \tag{5-1}$$

where

---

[6] $(x_1\!: 0, 1)$ is used to represent the set of all liquid mole fractions x_1 satisfying $0 \le x_1 \le 1$. Similar notation appearing in the text is read in-kind.

n : is the number of data points
$P_{exp,i}$: is the experimental total pressure (mmHg) of the i^{th} data point
$P_{calc,i}$: is the calculated total pressure (mmHg) of the i^{th} data point

The ratios of the objective function values of UNIFAC fittings to the component-based methods of Wilson, NRTL, and UNIQUAC were then calculated. Since deviations of 0.001 are often acceptable in practical applications (i.e. sufficiently precise), a minimum relative deviation of 10^{-6} was used when computing the resulting ratios. Therefore, any fitness values less than 10^{-6} obtained from the regression of a dataset using Equation (5-1) were set equal to the minimum value of 10^{-6} in order to get meaningful and comparable results.

5.1.2.1. *Calculations*

Although models with more adjustable parameters typically yield better results, the added complexity may lead to instabilities within the equation (Rarey, 2005). Figure 10 is used to illustrate this point, where some arbitrary data (linear in nature) has been fitted with a linear equation-form and a high-ordered polynomial (6 degrees in this case). Although the high-ordered polynomial fits the regressed data nearly perfectly, it does not follow the overall linear trend (or behaviour) of the data it was fitted to; the high-ordered polynomial is said to “over fit” the data. On the other hand the linear equation correctly captures the character/trend of the data, and thus could be used to interpolate and extrapolate between and outside of the data points it was fitted to (i.e. it provides a sufficiently precise description of the behaviour represented by the data). The polynomial, however, is just too flexible (has too many degrees of freedom) and would likely result in large errors whenever points outside those in which it was fitted to are calculated (i.e. the trend of the data is falsely captured).

In a likewise fashion it becomes necessary to limit the number of maingroups that are fitted for each dataset so that the results are comparable to those obtained from the Wilson, NRTL, and UNIQUAC equations. Since the third adjustable parameter of NRTL often contributes negligibly to the overall fit of the model (except in cases of limited miscibility) the UNIFAC fittings were initially restricted to systems composed of 2 different maingroups, but this was later extended to binary mixtures consisting of 3 and 4 maingroups for completeness/comparison.

In the fitting of each dataset with UNIFAC, only datasets which had group fractions in the pure components different from the group fractions of the mixture were used. This discarded mixtures of isomers, and ensured that the total excess Gibbs free energy of the mixture could be determined via group approach (i.e. the residual contributions would not be zero, and thus the binary interaction parameters between groups could be determined/fitted).

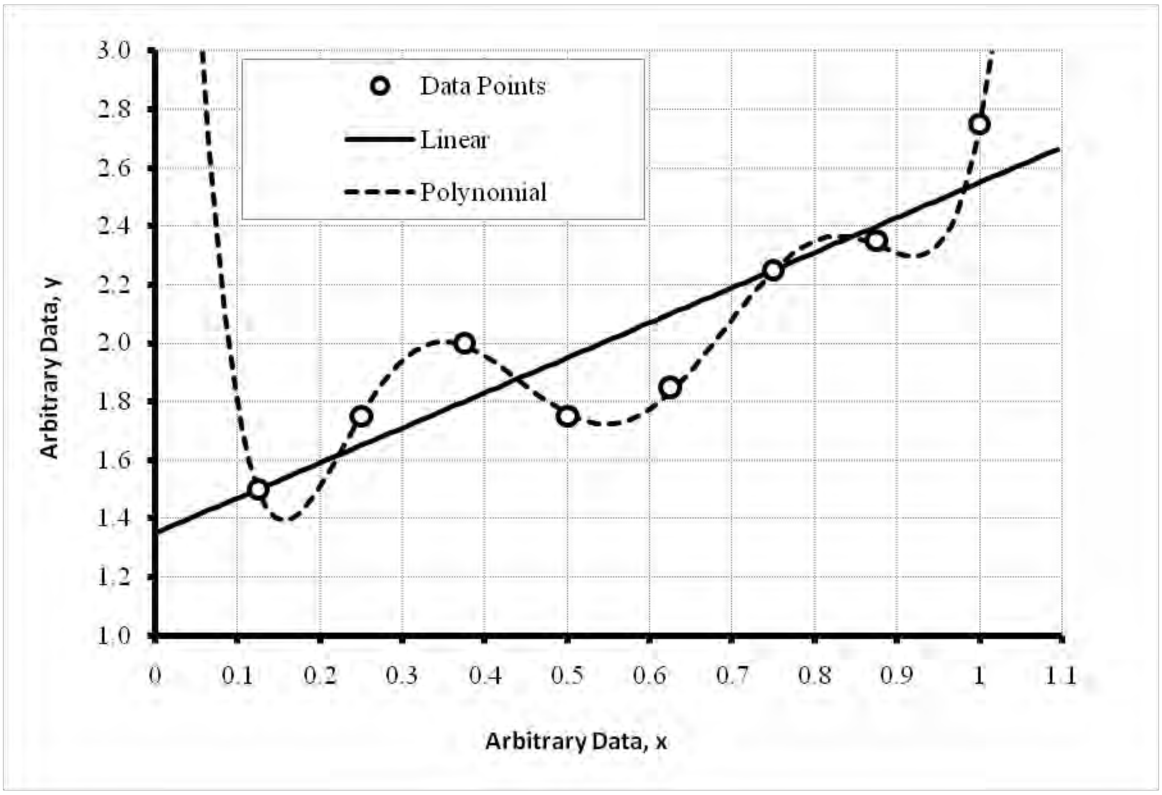


Figure 10 An illustration of over fitting where some arbitrary data (linear in character) was fitted by a linear equation and a high-ordered 6th degree polynomial.

The regression results using the UNIFAC method are compared against the results using the UNIQUAC equation in Figure 11, against the regression results using the NRTL equation in Figure 12, and against the regression results using the Wilson equation in Figure 13. Each data point of each line represents the resulting objective function value obtained using the UNIFAC method (fitted) divided by the objective function result obtained from each of the respective component-based models. The $\log_{10}$ of these ratios are depicted on the ordinates of each graph, and are ordered from smallest to largest in order to obtain a continuous line for each comparison. Any line drawn perpendicular from the abscissa then represents the percentage of datasets that have a ratio equal to or less than the value on the line it intersects at that point.

For comparison, the results have been grouped according to the number of maingroup interactions that were fitted using the UNIFAC method; since the abscissa is reported in percentage of datasets, the number of datasets having ratios equal to or less than a point on a particular line can be computed using Table 1 below.

Number of Maingroups	Number of Binary Isothermal Datasets, xP(T)
2 Maingroups	959
3 Maingroups	2,231
4 Maingroups	467
Total Number of Datasets	**3, 657**

Table 1 Number of datasets taken from the DDB-VLE containing 2, 3, or 4 maingroups from the DDB (DDBST Software and Separation Technology GmbH, 2009).

Since the ordinates of Figure 11 through Figure 13 are of $\log_{10}$, any values less than y = 0 represent cases in which the group contribution approach (using UNIFAC) fitted the datasets better than the component-based approaches of UNIQUAC, NRTL, or Wilson respectively. In either graph the term “crossover” is used to represent the point at which each line crosses *over* the y = 0 line (the percentage of data that is represented more accurately or equivalently using the group contribution approach of UNIFAC), and the term “midpoint” represents the point value (the ratio) at 50% of the datasets for each line.

Looking at the crossover and midpoints of each of the lines gives an indication of how well the group contribution approach using UNIFAC stacks up against the component-based models of UNIQUAC, NRTL, and Wilson; these points are summarized below, where the log-base$_{10}$ values and the absolute values (in parentheses) for the midpoints are both shown for convenience:

Number of Maingroups	Crossover (Figure 11) - UNIQUAC	Crossover (Figure 12) - NRTL	Crossover (Figure 13) - Wilson	Midpoint (Figure 11) - UNIQUAC	Midpoint (Figure 12) - NRTL	Midpoint (Figure 13) - Wilson
2 Maingroups	43.7%	35.8%	23.5%	0.0023 (1.0052)	0.0611 (1.1511)	0.2168 (1.6473)
3 Maingroups	88.4%	78.1%	76.2%	-0.0854 (0.8215)	-0.1106 (0.7751)	-0.0639 (0.8632)
4 Maingroups	84.4%	80.5%	88.4%	-0.1310 (0.7396)	-0.1537 (0.7019)	-0.1738 (0.6702)

Table 2 Percentage of datasets at the crossover points, and the midpoint values for each line in Figure 11 - Figure 13. The crossover point represents the point at which the ratio of the Pdev^2 values of UNIFAC to the Pdev^2 values of the component-based methods is equal to 1 (or equal to 0 on a $\log_{10}$ scale as shown in the figures). The $\log_{10}$ and absolute values of the ratios at 50% of the datasets for each line (the values at the midpoints) are also shown, where the $\log_{10}$ values are those in parentheses.

As can be seen from the Table 2 above and Figure 11 through Figure 13 below, the component-based models of UNIQUAC, NRTL, and Wilson outperformed the direct fitting of the UNIFAC method for binary mixtures represented by two maingroups. As mentioned earlier these results (binary mixtures which only require the binary interactions between two maingroups to be calculated via UNIFAC) likely provide the best results for the comparison of the two methodologies. In this case the results obtained from the component-based models show improvement over the group contribution approach using the UNIFAC method, where the results obtained using the UNIQUAC equation are better or equivalent 56.3% of the time (with a crossover point of 43.7%). Those obtained using the NRTL equation are better or equivalent 64.2% of the time (with a crossover point of 35.8%), and those obtained using the Wilson equation are better or equivalent than the direct fitting results using the UNIFAC method 76.5% of the time (with a crossover point of 23.5%). Not only do the low crossover values show that there is no overwhelming benefit to using the group contribution approach over the component-based approach (when limiting the discussion to cases involving only two maingroups), but the relatively high midpoint values show that the results are not even close to being similar between the two methodologies.

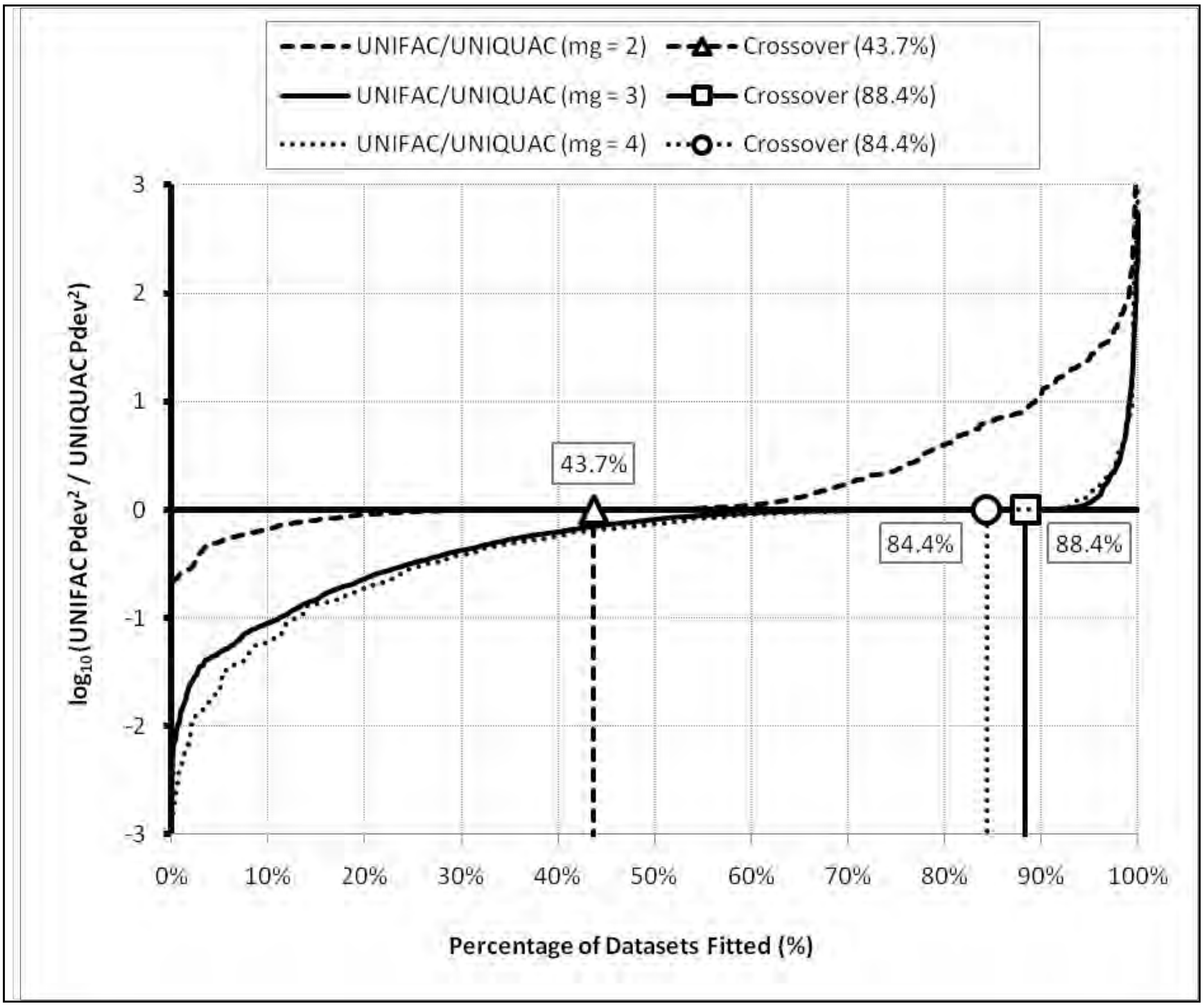


Figure 11 Log$_{10}$ ratio of Pdev^2 values calculated using Equation (5-1), where the results using the UNIFAC method with fitted maingroup parameters on a case-by-case basis (numerator) are divided by the results of those same datasets using the UNIQUAC equation (denominator).

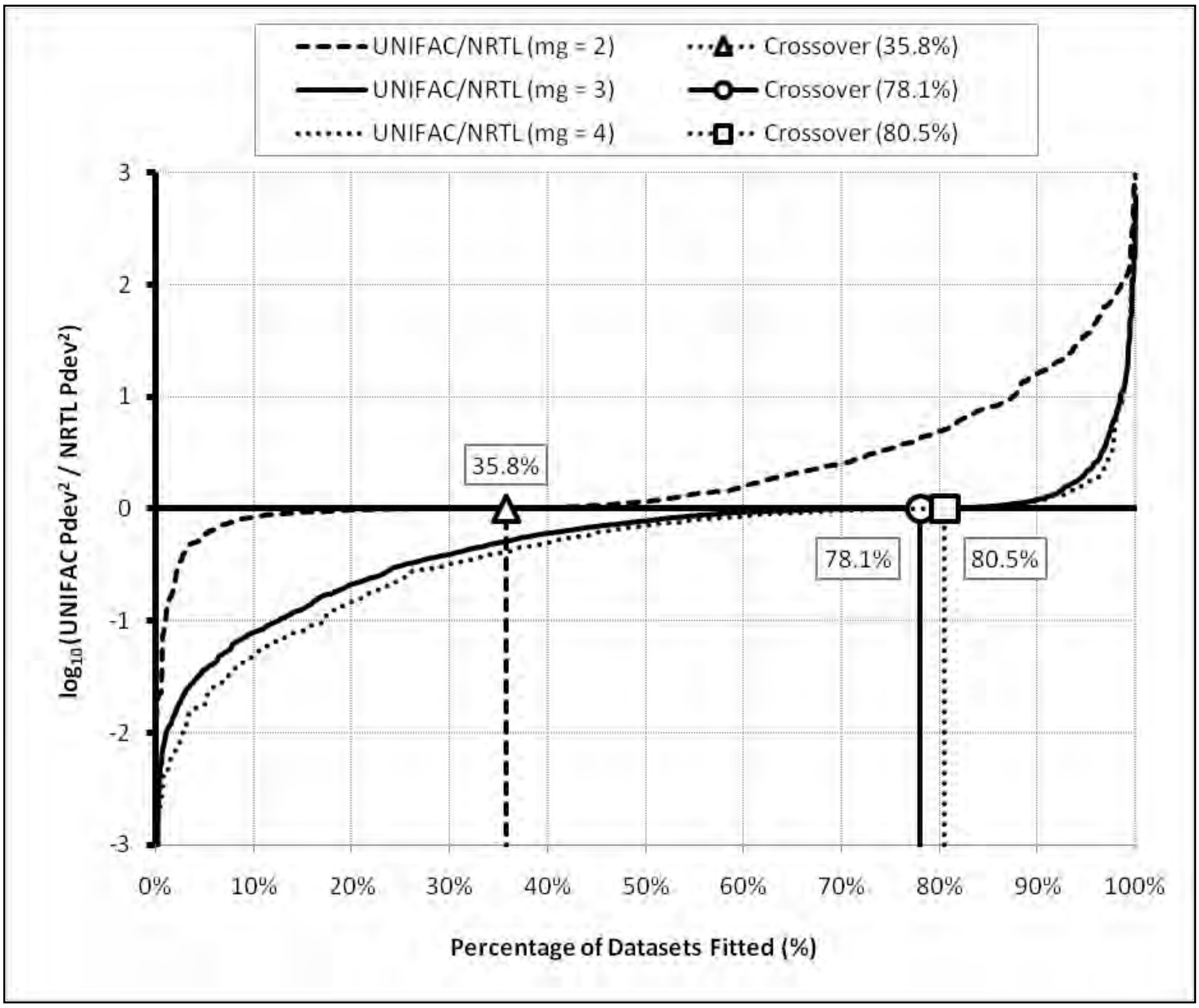


Figure 12 Log_{10} ratio of Pdev^2 values calculated using Equation (5-1), where the results using the UNIFAC method with fitted maingroup parameters on a case-by-case basis (numerator) are divided by the results of those same datasets using the NRTL equation (denominator).

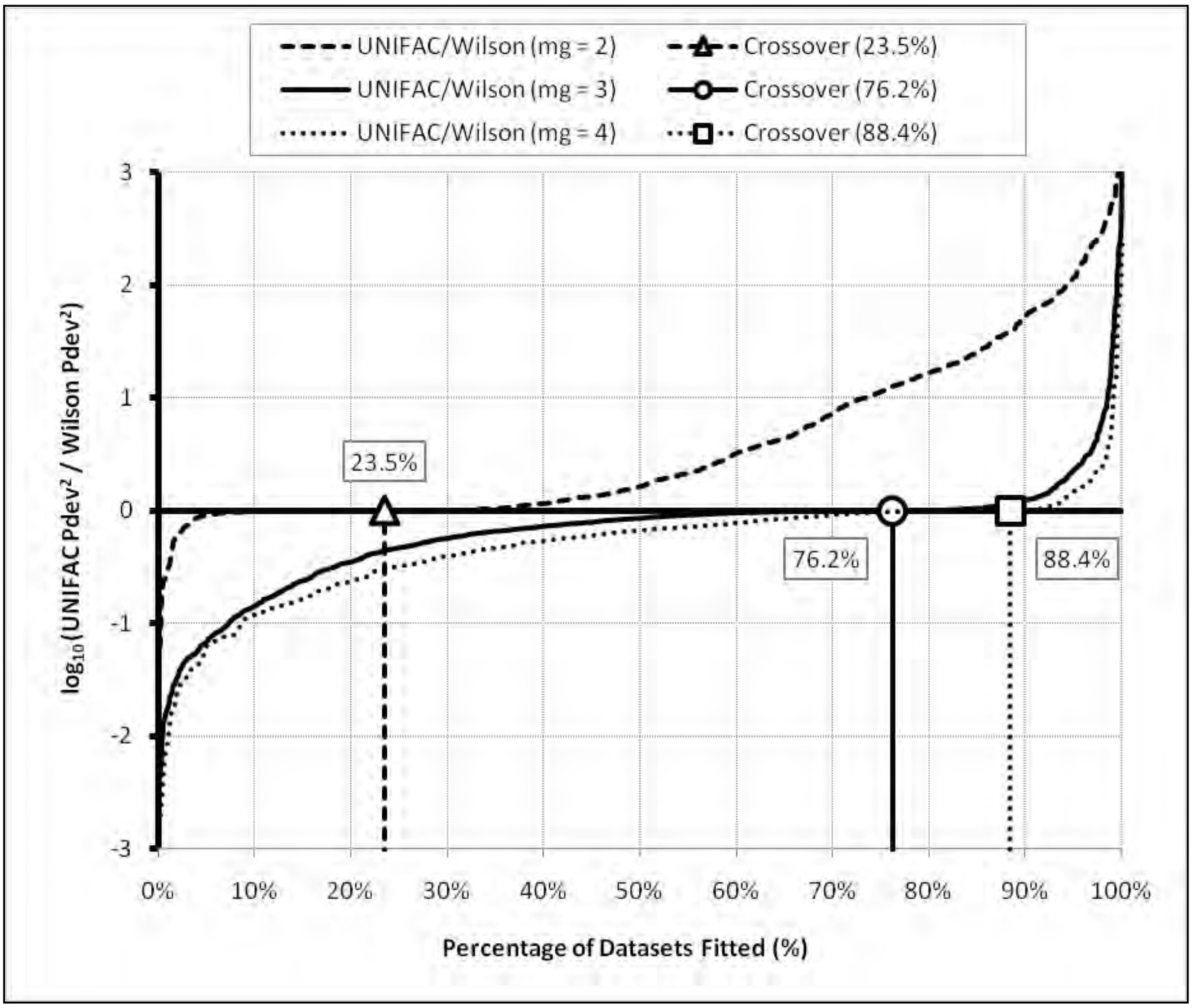
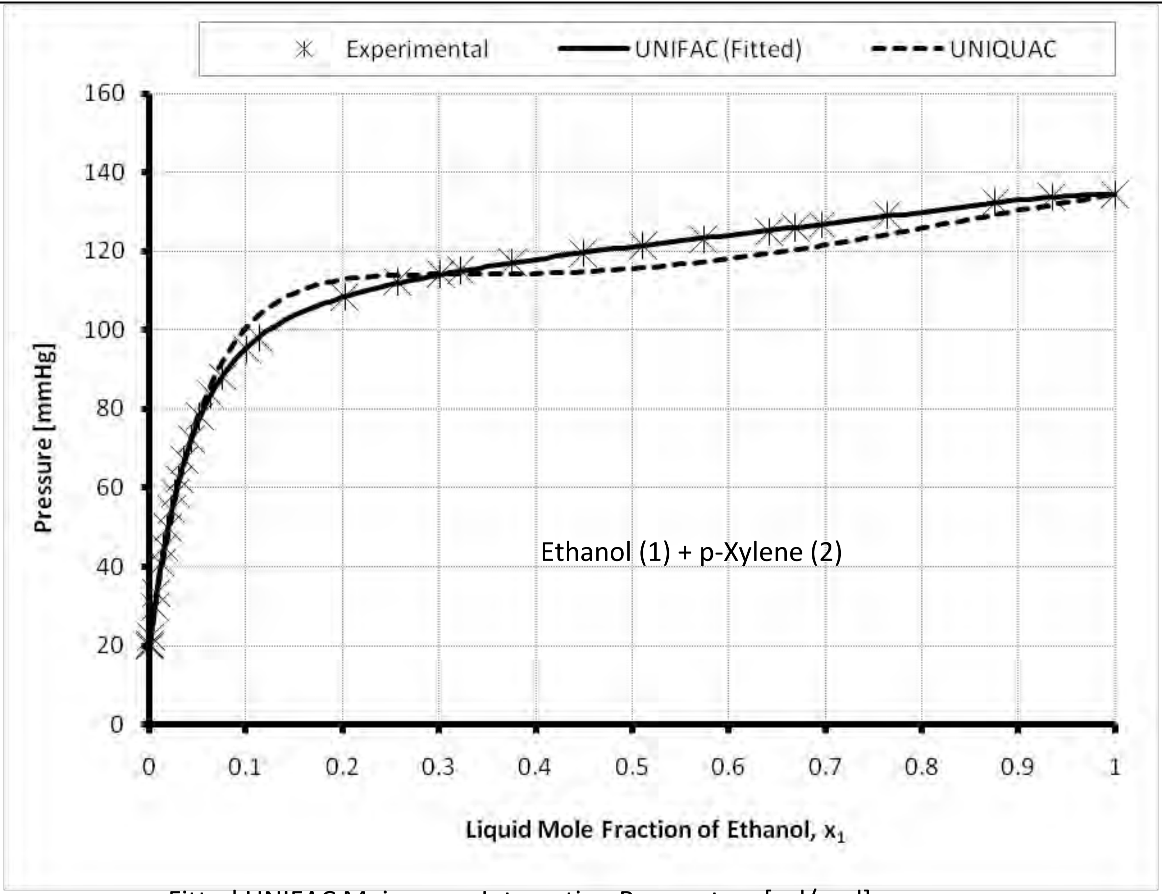


Figure 13 Log_{10} ratio of Pdev^2 values calculated using Equation (5-1), where the results using the UNIFAC method with fitted parameters on a case-by-case basis (numerator) are divided by the results of those same datasets using the Wilson equation (denominator).

As the number of maingroups fitted directly using the UNIFAC method is increased, however, the fitting results favour the group contribution approach over that of the component-based approach using the UNIQUAC, NRTL, and Wilson equations (i.e. for experimental datasets requiring more than two maingroups in order to be calculated using the UNIFAC method). Given the increase in the number of parameters being fitted (i.e. an increase in complexity) this result is not wholly unexpected, but what is interesting is the level of improvement that is gained when the number of maingroup parameters being fitted is increased from two to six. The crossover point increases by a factor of 1.9 (84.4%/43.7%, see Figure 11) compared to UNIQUAC, a factor of 2.2 (78.1%/35.8%, see Figure 12) compared to NRTL, and by a factor of 3.2 (76.2%/23.5%, see Figure 13) compared to Wilson. Further improvements, although marginal in comparison, were obtained when the number of parameters being fitted was increased from six to twelve. Figure 14 is used to illustrate typical results for these cases, where the difference in the number of adjustable parameters contributes to the differences between the fitting results of each method.



Fitted UNIFAC Maingroup Interaction Parameters [cal/mol]:

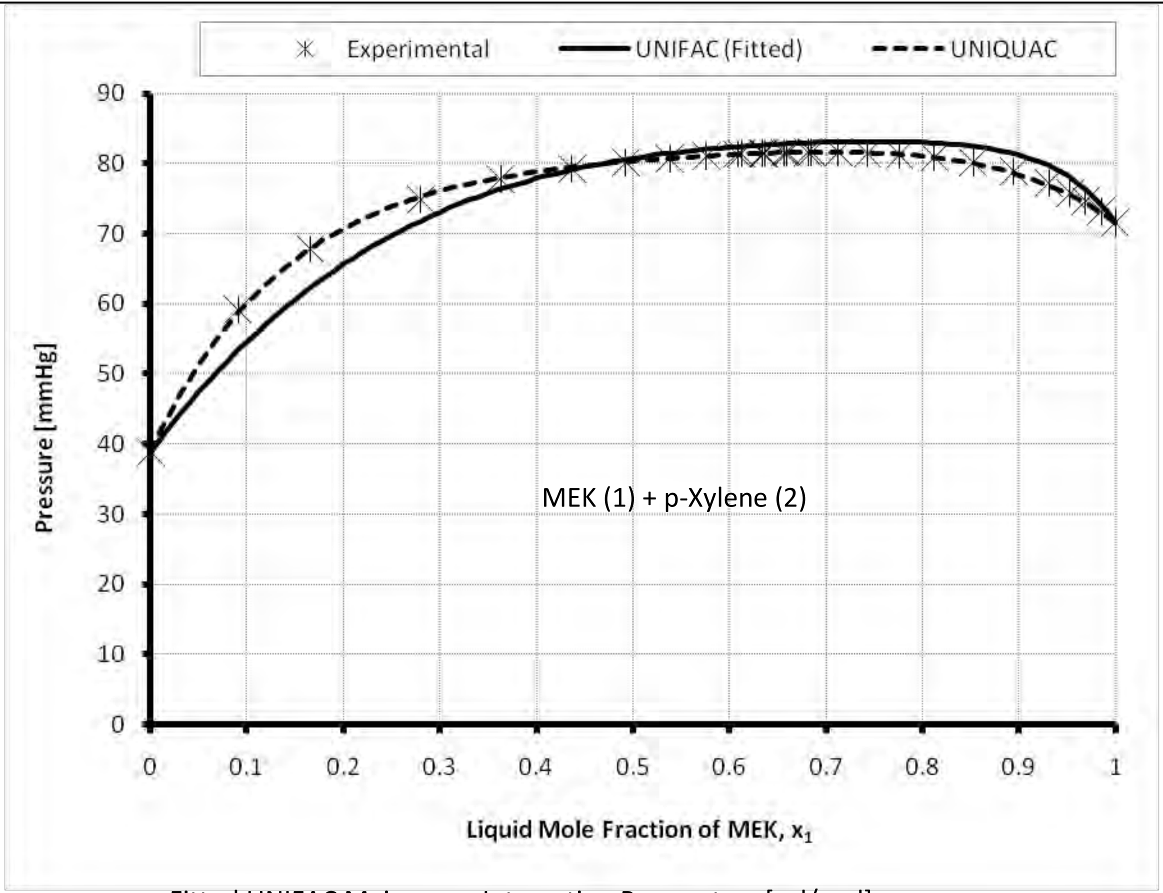
Maingroup	**1**	**3**	**4**	**5**
1	0	631.993	128.292	-464.786
3	1040.322	0	138.379	390.602
4	207.660	-296.777	0	158.473
5	-379.484	-307.865	-481.701	0

Fitted UNIQUAC Interaction Parameters [cal/mol]:

Parameter	**Value [cal/mol]**
τ_{12}	-234.017
τ_{21}	1120.603

Figure 14 Px(T) dataset of ethanol (1) and p-xylene (2) at 313.15 K (DDB-VLE set number 12637). Experimental pressures are compared against calculated results obtained using UNIQUAC and the UNIFAC method, where both models were fitted directly to the experimental data. The fitted parameters for UNIQUAC and UNIFAC are also provided.

When the complexity of the group contribution approach using the UNIFAC method is limited to the comparable case of fitting binary interaction parameters between only two maingroups, however, the advantage of representing the behaviour of the real mixtures tested goes to the component-based approach. This observation is maintained even if the discussion is limited to a comparison between the fitting results of UNIFAC and UNIQUAC, which represents the closest comparison between the two methodologies, since the UNIFAC method is based on the UNIQUAC equation. Figure 15 is used to illustrate this point for the system MEK (1) + p-xylene (2), where the UNIFAC method fits the experimental data with less precision than the UNIQUAC equation. Since this system only requires the fitting of two parameters to the experimental data for these methods, the differences between the fitting results then represent how each method treats the concentration dependence of the activity coefficients for the mixture. Therefore instead of making adjustments directly to the interaction parameters between maingroups, using a component-based model to correct for the predictions made using a group contribution method seems more prudent (an easier alternative).



Fitted UNIFAC Maingroup Interaction Parameters [cal/mol]:

Maingroup	1	9
1	0	253.2095
9	2290.575	0

Fitted UNIQUAC Interaction Parameters [cal/mol]:

Parameter	Value [cal/mol]
τ_{12}	-135.499
τ_{21}	490.425

Figure 15 Px(T) dataset of methyl ethyl ketone (MEK) (1) and p-xylene (2) at 293.15 Kelvin (DDB-VLE set number 13795). Experimental pressures are compared against calculated results obtained using UNIQUAC and the UNIFAC method, where both models were fitted directly to the experimental data (both only requiring two fitted parameters each). The fitted parameters for UNIQUAC and UNIFAC are also provided.

One explanation, for the differences that exist between the two approaches, is that the group approach incorrectly represents the concentration dependence of the activity coefficients. While the component-based methods of Wilson, NRTL, and UNIQUAC use liquid mole fractions as their primary concentration variable, the group contribution methods such as UNIFAC use “group fractions” (see Equation (3-50)). This means, for instance, that the group-based methods will cover a different concentration range than component-based methods, and they may not cover the full concentration range between 0 and 1 (see Figure 16). This may in part explain why the component-based methods proved superior in fitting the data, for the comparable cases where the component system is represented by two UNIFAC maingroups (see Figure 11 through Figure 13).

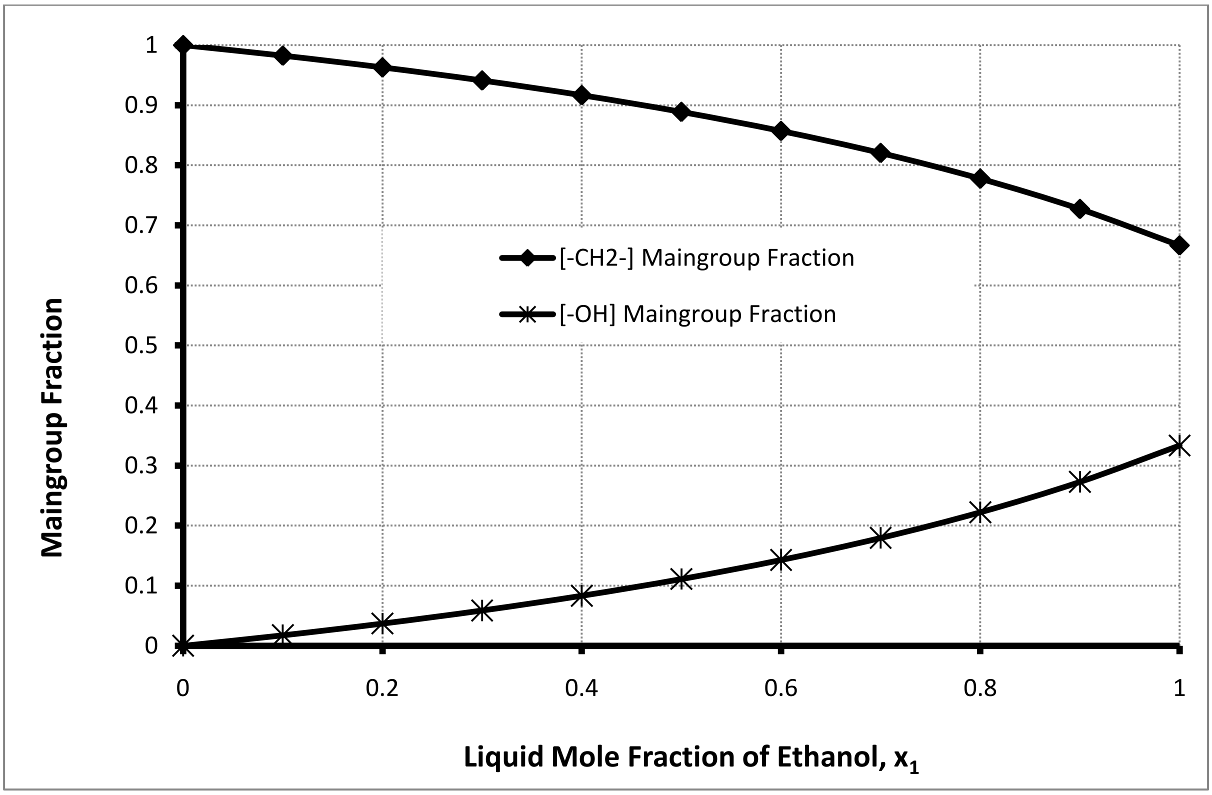
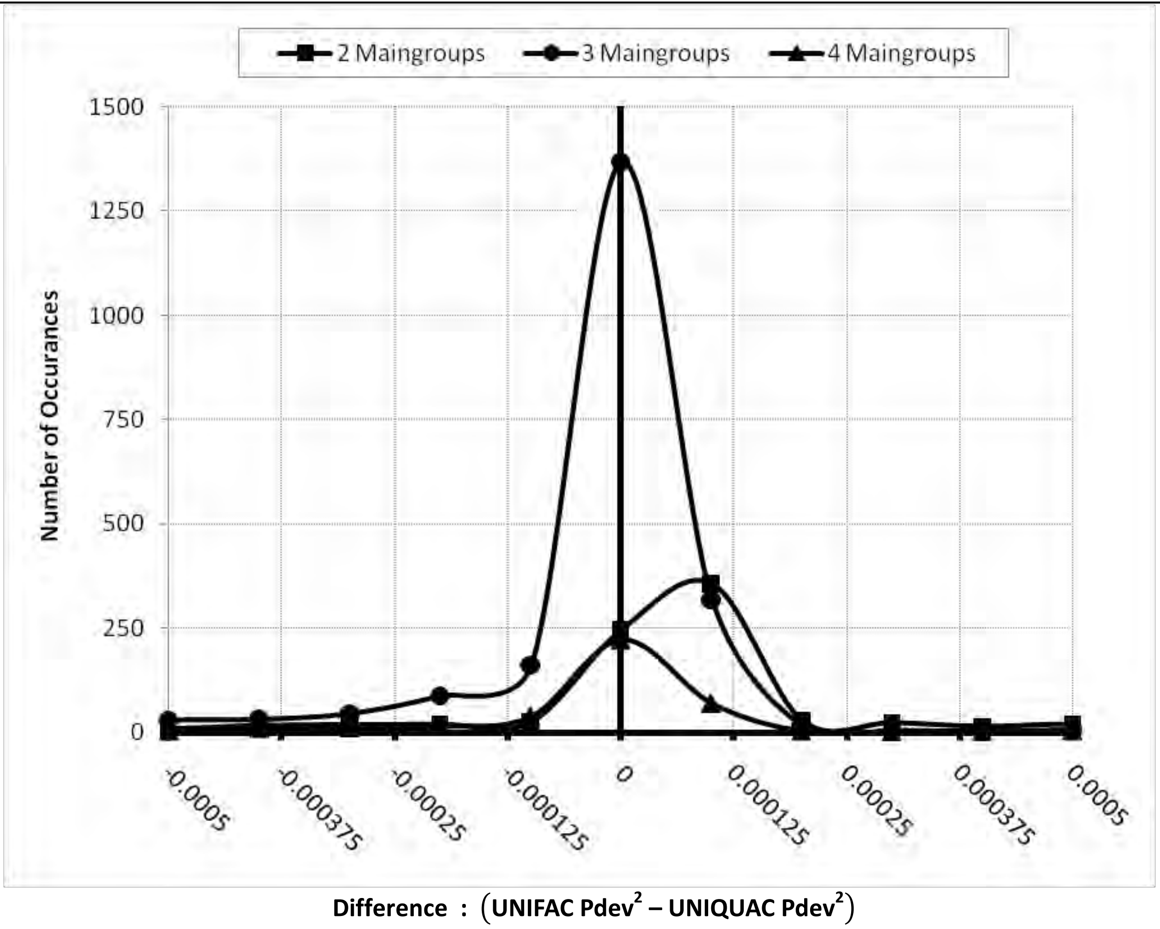


Figure 16 The maingroup fractions of CH2 and OH plotted as a function of ethanol liquid mole fractions, for the system ethanol (1) / n-hexane (2).

It should be noted, however, that the differences between the fitting results using UNIFAC and those using the component-based methods of UNIQUAC, NRTL, and Wilson are in large part small. One should keep in mind that Figure 11 through Figure 13 mainly show which method is better than the other one is, and does not provide a good indication of the magnitude of the differences between the fitting results. As shown in Figure 17, for example, the differences between the fitting results of using the UNIFAC method and those using the UNIQUAC equation are often very similar to each other. For the comparable case of 2 maingroups, for instance, the average and maximum absolute deviations are only 7.569E-04 and 2.822E-04 respectively.



Data Series	Average Absolute Difference | Δ |	Maximum Absolute Difference	Number of Data Points
2 Maingroups	7.569E-04	4.087E-02	959
3 Maingroups	2.822E-04	3.827E-02	2231
4 Maingroups	6.892E-04	8.713E-02	467

Figure 17 A distribution of the differences between the fitting results (i.e. the Pdev^2 values) using the UNIFAC method and those obtained using the UNIQUAC equation. The average and maximum absolute differences are also provided; furthermore, these curves correspond to those shown in Figure 11.

5.2. Ability to Predict Activity Coefficients at Infinite Dilution

In the preceding chapter, the component-based and group contribution methodologies were compared. The advantage of representing the real behaviour of binary mixtures was shown to lie with the component-based approach using the Wilson, NRTL, and UNIQUAC equations, when compared to the direct fitting of binary interaction parameters between two maingroups using the UNIFAC method (i.e. comparable to the fitting of component-based model parameters). Thus, whichever component-based model is used within the proposed hybrid model will be responsible for correcting any deviations that may exist between the predictive results (coming from the group contribution layer) and available experimental data. The more accurate the predictive model is, the less the component-based layer/model will have to compensate for; accordingly it becomes of interest to test the ability of the group contribution models to correctly represent the behaviour of very dilute binary mixtures, where extreme changes in property values are often observed. Not only is this study relevant to the proposed hybrid method, but the results should be of general interest to practitioners.

In Industry, predictive methods such as the UNIFAC method, are often used to supplement the working knowledge of how real multicomponent mixtures behave, where the correct representation of activity coefficients at infinite dilution (γ_i^∞) is often important for a suitable representation of many separation processes (like distillation columns for instance). It is often the behaviour in this region that determines how difficult a separation will be, the size of a column's stripping section, the operating conditions required to obtain a targeted purity, and for a large part whether or not the separation is even feasible and/or practical to perform.

In the results and discussion that follows, the experimental γ^∞ values of the ACT database of the Dortmund Data Bank (DDBST Software and Separation Technology GmbH, 2009) were used for the comparison of the predictive methods UNIFAC and mod. UNIFAC (Do), and their MRR variations (via direct substitution of the combinatorial expression, Equation (4-1) through Equation (4-6)). As in the subsequent Chapters, the Consortium model parameters (2008) were used. In order to remove datasets of insufficient quality, the following restrictions were used:

> The ability of the UNIFAC and mod. UNIFAC (Do.) methods to predict the solubilities of hydrocarbons in water is a widely known and accepted model weakness (Banerjee, 1985) (DDBST Software and Separation Technology GmbH, 2009). Hydrocarbon-water systems were therefore omitted from the study.[7]
>
> Experimental results obtained by liquid-liquid-chromatography (LLC) were also omitted, since these measurements are largely inconsistent with more accurate measurement methods available (Sherman, et al., 1996).
>
> The database was further limited to γ^∞ values of ≤ 500, since large discrepancies between authors typically arise for values > 500 (Mu, et al., 2007).

The restrictions resulted in a final database consisting of 46 858 experimental γ^∞ values from 15 663 unique binary systems (from the original 55 040 datasets, and 16 934 unique binary systems within

---

[7] See (DDBST Software and Separation Technology GmbH, 2009) for an empirical correction for representing hydrocarbon solubilities.

the ACT database); a breakdown according to temperature and measurement method is shown in Figure 18 below.

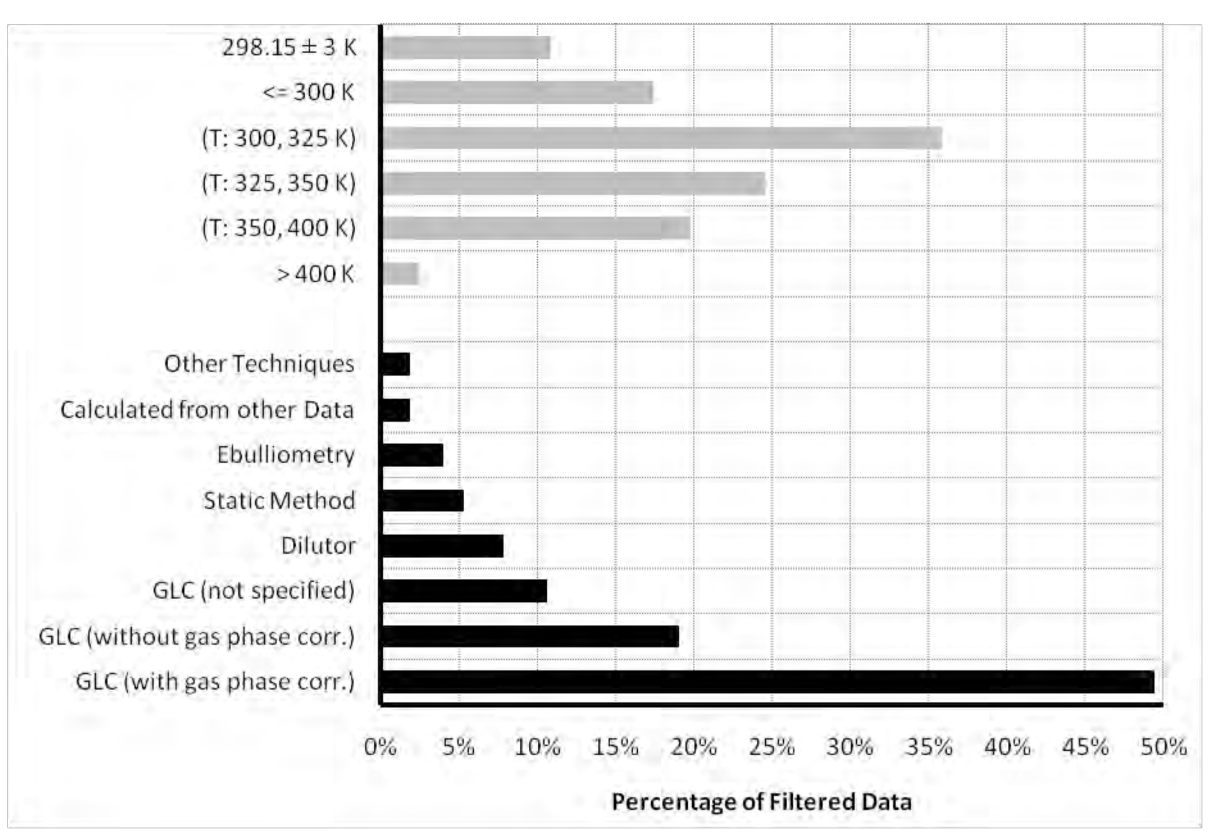


Figure 18 Breakdown of the 46,858 datasets from the ACT database of the Dortmund Data Bank (DDBST Software and Separation Technology GmbH, 2009), showing the distribution of datasets according to temperature ranges (top) and measurement methods (bottom).

As mentioned in Chapter 4.1, recent work on the predictions of γ^{∞} values in hydrocarbon solvents performed by Moller (2010) has resulted in a modified GK-FV free volume expression (MRR, Equation (4-1) through Equation (4-6)). The MRR combinatorial was found to provide a much-improved prediction for large solutes in small solvents, while maintaining near-equivalent results for small solutes in larger solvents. This expression, however, was developed and tested for its ability to predict/extrapolate limiting component solubilities in saturated hydrocarbons only, and not for the broader ability of predicting limiting activity coefficients for various components in *any* solvent.

For this reason variations of the UNIFAC and mod. UNIFAC (Do.) methods were also tested via the direct substitution of the MRR combinatorial, without the fitting of any new GC model parameters; for convenience, these variations will be referred to as UNIFAC-MRR and mod. UNIFAC (Do.)-MRR respectively.

Moller (2010) also observed, however, that the fitted volume and surface area values of the mod. UNIFAC (Do.) method may result in incorrect combinatorial values; therefore in the implementation of mod. UNIFAC (Do.)-MRR, the volume and surface area values were determined from the method of Bondi (1964), since these values are physically more realistic (i.e. not just additional parameters, which were, fitted with the group interaction parameters of the model). This means in cases like alkane/alkane mixtures, where binary interactions between molecular groups will be identical (i.e. the residual contribution will be zero) and the calculated results of the UNIFAC-MRR and mod. UNIFAC (Do.)-MRR methods will be identical.

5.2.1. Calculations[8]

The activity coefficients at infinite dilution of binary systems were predicted with the group contribution equations of UNIFAC and mod. UNIFAC (Do.), including the variations obtained by direct substitution of the MRR combinatorial recently developed (without the fitting of any new group contribution parameters). For comparison, the relative errors were calculated for each dataset using the following equation:

$$\Delta\gamma^{\infty}_{R_{err}} = \frac{\gamma^{\infty}_{calc} - \gamma^{\infty}_{exp}}{\gamma^{\infty}_{exp}} \tag{5-2}$$

where

γ^{∞}_{exp} : is the experimental activity coefficient at infinite dilution of the solute
γ^{∞}_{calc} : is the calculated activity coefficient at infinite dilution of the solute

Distribution curves were then determined from the relative error values that were calculated for each group contribution model tested. The resulting distributions are presented in Figure 19 and Figure 20, where the abscissa represents the relative error bins $(\mathrm{R_{err}}\colon -1, 1)$ used to generate the distribution curves for the methods. The ordinates in these cases represent the numbers of occurrences (or datasets in the case of DDB-ACT) which were predicted to have the same relative error as an intersecting perpendicular line drawn from the abscissa. “Mid-point” line values are also included for each distribution curve in Table 3, and represent the median of their respective distribution curves (determined for each line via the area under the curve). Included in Table 3 is the standard deviation of the relative errors used to generate the distribution curves, and provides a relative measure of the width/breadth of each of these distributions against one another.

It should be noted, however, that the mid-point lines for the distribution curves are calculated using a higher resolution than what is depicted in Figure 19 and Figure 21. Where the depicted distribution curves were calculated using intervals of 10% in relative errors (of activity coefficient values at infinite dilution), the mid-point values listed in Table 3 and Table 4 were determined using a resolution of 0.1% in relative errors. This was done to enhance the readability of the results, and should not greatly sacrifice the ability to make *visual* comparisons via Figure 19 and Figure 21. The deviations depicted in Figure 19 and Figure 21 are still comparable and give an indication of the “relative” adequacy of each equation as compared against each other.

---

[8] UNIFAC Consortium values (2008) were used to produce the results, unless specifically stated otherwise.

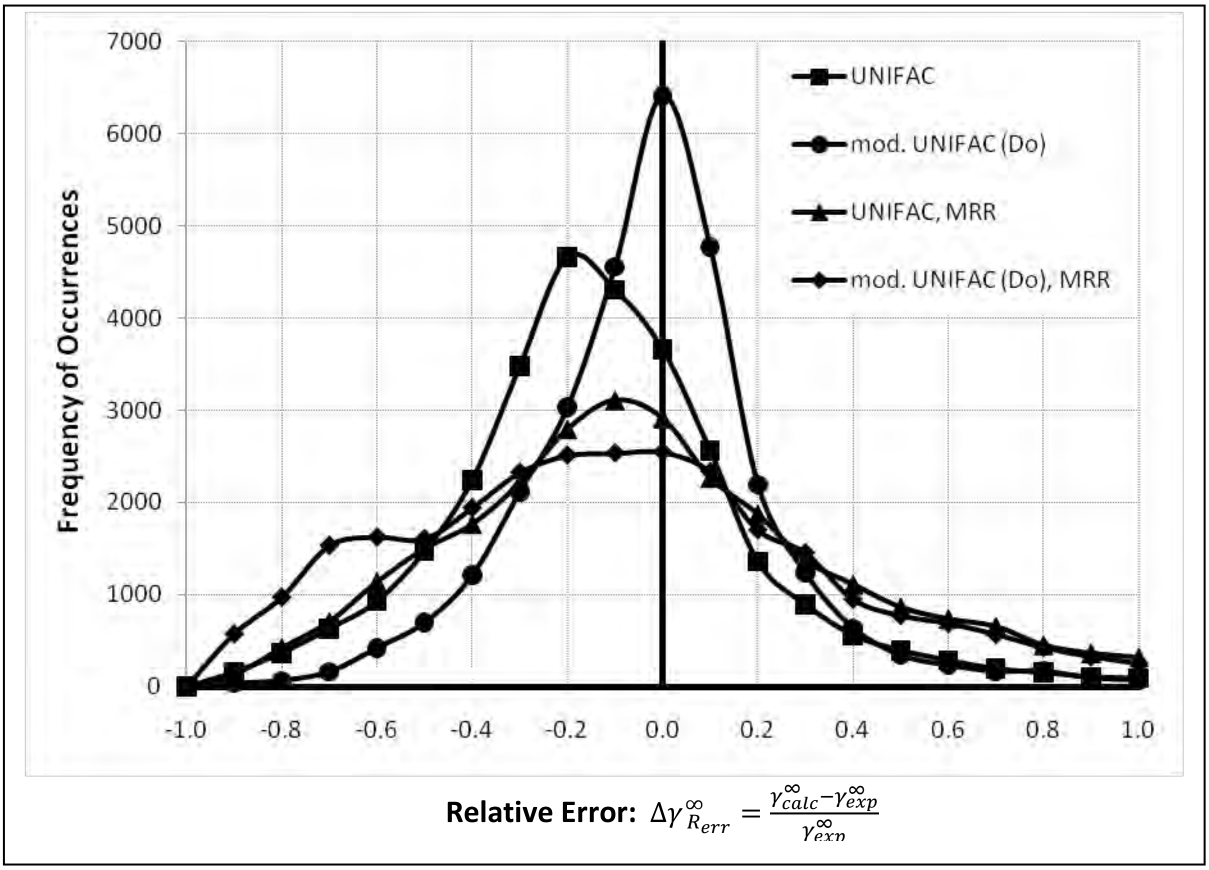


Figure 19 Distribution curves of the relative errors in predicting activity coefficients at infinite dilution; obtained from predicting DDB-ACT data using UNIFAC and mod. UNIFAC (Do.) methods (2008 Consortium parameters), including the variations obtained by directly replacing the combinatorial expressions in each equation with the MRR combinatorial.

Method	Mid-point Values for the Relative Error Distributions	Percentage of Datasets Under Predicted	Variance Values for the Relative Error Distributions
UNIFAC	-19.2 %	74.9 %	9.2 %
UNIFAC-MRR	-11.2 %	64.0 %	15.7 %
mod. UNIFAC (Do.)	-6.5 %	57.1 %	6.4 %
mod. UNIFAC (Do.)-MRR	-17.0 %	62.2 %	18.2 %

Table 3 Mid-point values and the variance for the distribution curves of relative error from Figure 19 are tabulated, along with the percentage of calculated $\gamma^{\infty}_{i,calc}$ values smaller than the experimental values $\gamma^{\infty}_{i,exp}$ for each method.

The original UNIFAC method significantly under predicts the real behaviour at infinite dilution when compared to the mod. UNIFAC (Do.) method (see Figure 19). The distribution of relative errors using UNIFAC results in a mid-point value of -19.2% in relative error (with a variance of 9.2%), and the mod. UNIFAC (Do.) method results in a mid-point value of -6.5% in relative error (with a variance of 15.7%). The differences between the two distribution curves then represent the differences between the methods of UNIFAC and mod. UNIFAC (Do.).

The use of temperature-independent parameters in the UNIFAC method limits the types of physical properties used for the fitting of the binary interaction parameters between maingroups, and likely contributes to the large negative deviations in relative error produced by the method.

Additional error is also introduced by the combinatorial expressions of UNIFAC and mod. UNIFAC (Do.), where large errors are observed for large in small molecules (Moller, 2010).

The UNIFAC method is mainly fitted to VLE data within a limited concentration range (Gmehling, 1999), and sometimes limiting activity coefficients at infinite dilution (γ_i^∞) are used (Wittig, et al., 2003). This essentially means that any prediction within the dilute region is an extrapolation outside of the training set used to fit the model parameters—typically (x_i: $0.05, 0.95$). Given the extreme values and sharp slopes of the $\ln \gamma_i^\infty(x_i)$ curve in the diluted regions, the method, not too surprisingly, routinely under predicts the real behaviour of binary mixtures at infinite dilution (74.9% of the time, Table 3).

The mod. UNIFAC (Do.) method, on the other hand, uses temperature dependent parameters and additional physical properties for the fitting purposes (e.g. h^E, γ^∞, etc.), and includes a modified combinatorial expression to improve the predictions of activity coefficients at infinite dilution for asymmetric mixtures (Weidlich, et al., 1987). This means that mod. UNIFAC (Do.) contains knowledge of the entire composition range, versus the UNIFAC method. This essentially eliminates the “leap of faith” (the extrapolation to the dilute region) that the UNIFAC method is forced to make when predicting the real behaviour of binary mixtures at infinite dilution. This enables the mod. UNIFAC (Do.) method to realize roughly a 66% reduction in the relative error at the median of the curve, and about a 15% reduction in the percentage of datasets under predicted by the method (under predicted 64.0% of the time, Table 3). Given the use of temperature-dependent interaction parameters, the accuracy of the mod. UNIFAC (Do.) method is maintained across a large temperature-range as well (see Figure 20).

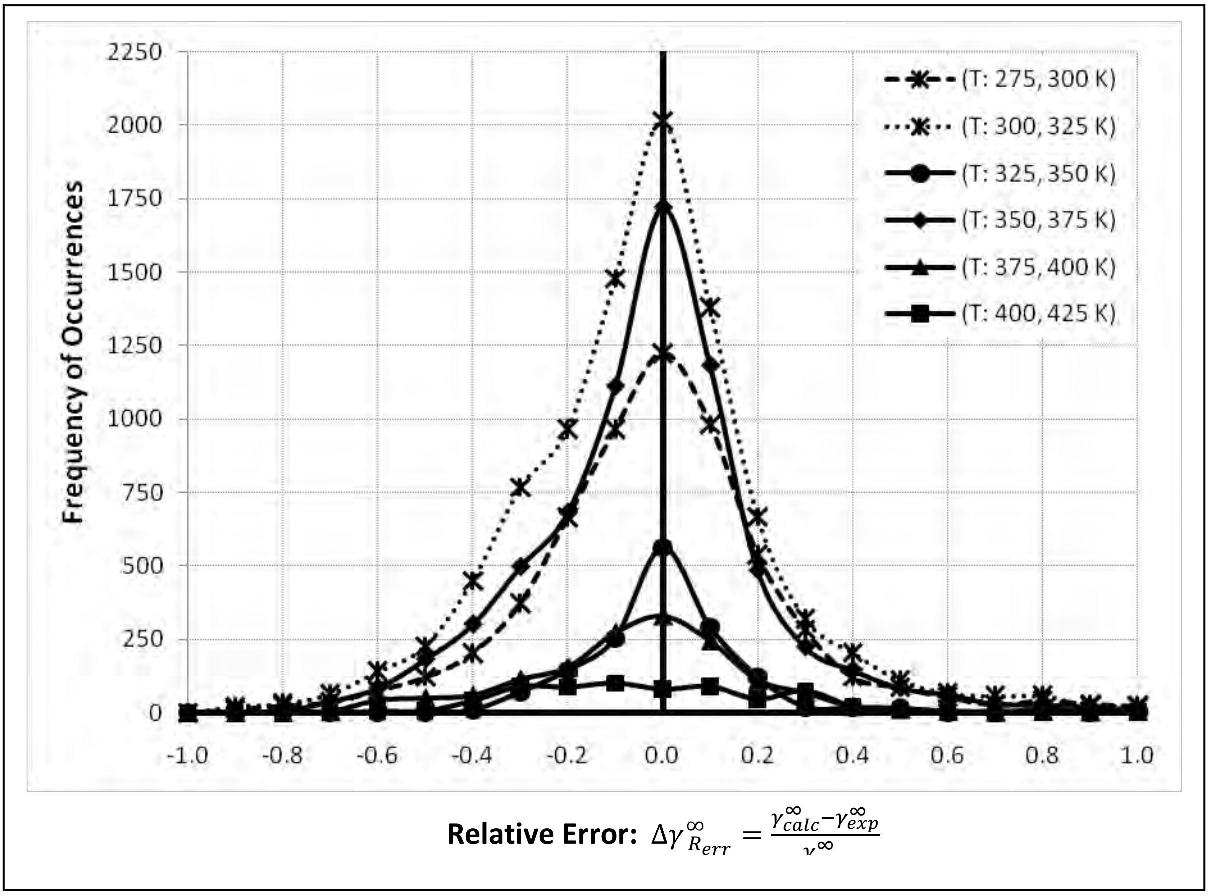


Figure 20 Distribution curves of the relative errors obtained from predicting DDB-ACT data using mod. UNIFAC (Do.) (2008 Consortium parameters), where the results have been grouped into temperature ranges incremented by 25 K (T: 275, 425 K).

Given the promising results of the new MRR combinatorial to predict the real behaviour of limiting component solubilities in hydrocarbon solvents, it has been included in the present study to see if these results are extended to other solvent-types. The MRR combinatorial was directly substituted into the UNIFAC and mod. UNIFAC (Do.) methods, without the fitting of any new model parameters. In the case of mod. UNIFAC (Do.)-MRR, the volume (r_i) and surface area (q_i) values were determined via the method of Bondi (1964). These method-variations were then used to predict the same DDB-ACT data; for comparison with the standard versions of the methods, the resulting distributions of the calculated relative errors were plotted on the same graph (Figure 19).

The characteristic curves of UNIFAC-MRR and mod. UNIFAC (Do.)-MRR then provides a visual representation of the effect that the MRR combinatorial has on the UNIFAC and mod. UNIFAC (Do.) methods (as a direct combinatorial replacement). The calculated medium-point for the UNIFAC-MRR method is -6.5% (compared to -19.2% for UNIFAC) and that of mod. UNIFAC (Do.)-MRR is calculated to be -17.0% (compared to -6.5% for mod. UNIFAC (Do.)). Although the direct substitution of the MRR combinatorial into UNIFAC results in a lower mid-point value, the overall effect results in a widening of the distribution curve (for the mod. UNIFAC (Do.)-MRR method as well). The UNIFAC-MRR method, for instance, results in roughly a 71% increase in the variance of the distribution of relative errors when compared to the standard version of UNIFAC ($71\% = 1 - 15.7\%/9.2\%$, see Table 3). This means that the absolute deviations and relative average deviations point-to-point will be higher than the standard versions of UNIFAC and mod. UNIFAC (Do.). The use of the MRR combinatorial, therefore, is limited to alkane-solvent systems only, and is not suitable as a direct replacement for either of the combinatorial expressions of UNIFAC or mod. UNIFAC (Do.).

In an effort to further elucidate the affects of the combinatorial expressions, however, the calculated distributions were further restricted to datasets consisting of saturated alkanes only. Since these alkanes are assembled from a set of common molecular subgroups (CH_3-, -CH_2-, >CH-, and >C<) belonging to the same maingroup (CH_2 maingroup of the UNIFAC and mod. UNIFAC (Do.) methods), the binary interactions between these molecular groups will be identical. The residual contribution, therefore, will be equal to zero, and the calculated activity coefficients for these mixture-types will be solely determined from the combinatorial contributions of each method (from the size and shape of the molecules alone). For binary systems, these combinatorial expressions can be written as the following:

$$ln\gamma_2^{C,\infty} = 1 - \frac{r_2}{r_1} + ln\left(\frac{r_2}{r_1}\right) - 5q_2\left[1 - \frac{r_2q_1}{q_2r_1} + ln\left(\frac{r_2q_1}{q_2r_1}\right)\right] \tag{5-3}$$

$$ln\gamma_2^{C,\infty} = 1 - \left(\frac{r_2}{r_1}\right)^{3/4} + ln\left(\frac{r_2}{r_1}\right)^{3/4} - 5q_2\left[1 - \frac{r_2q_1}{q_2r_1} + ln\left(\frac{r_2q_1}{q_2r_1}\right)\right] \tag{5-4}$$

$$ln\gamma_2^{C,\infty} = \left[\left(\frac{V_2^{FV}}{V_1^{FV}}\right) - \left(\frac{V_2^{FV\star}}{V_1^{FV\star}}\right)^{2/3}\right] * \left[1.15 - 0.15 * \exp(-0.5(q_1 - q_2)^2 - 0.5(r_1 - r_2)^2)\right] \tag{5-5}$$

These expressions are used to examine the Guggenheim-Staverman combinatorial used by UNIFAC (GS, Equation (5-3)), the modified GS expression of mod. UNIFAC (Do.) (MU, Equation (5-4)), and the MRR combinatorial expression (MRR, Equation (5-5)). In order to elucidate the effects of using the group volume and surface area parameters as additional fitting constants, the modified GS expression of the mod. UNIFAC (Do.) method was also calculated using the combinatorial parameters defined by UNIFAC for comparison (i.e. GS-3/4, Equation (5-5) with UNIFAC definitions). The resulting calculated distributions in relative error and mid-point lines for each expression (and option) tested are presented in Figure 21 for systems of saturated alkanes, where the mid-point values and percentage of datasets under predicted are tabulated in Table 4 for convenience.

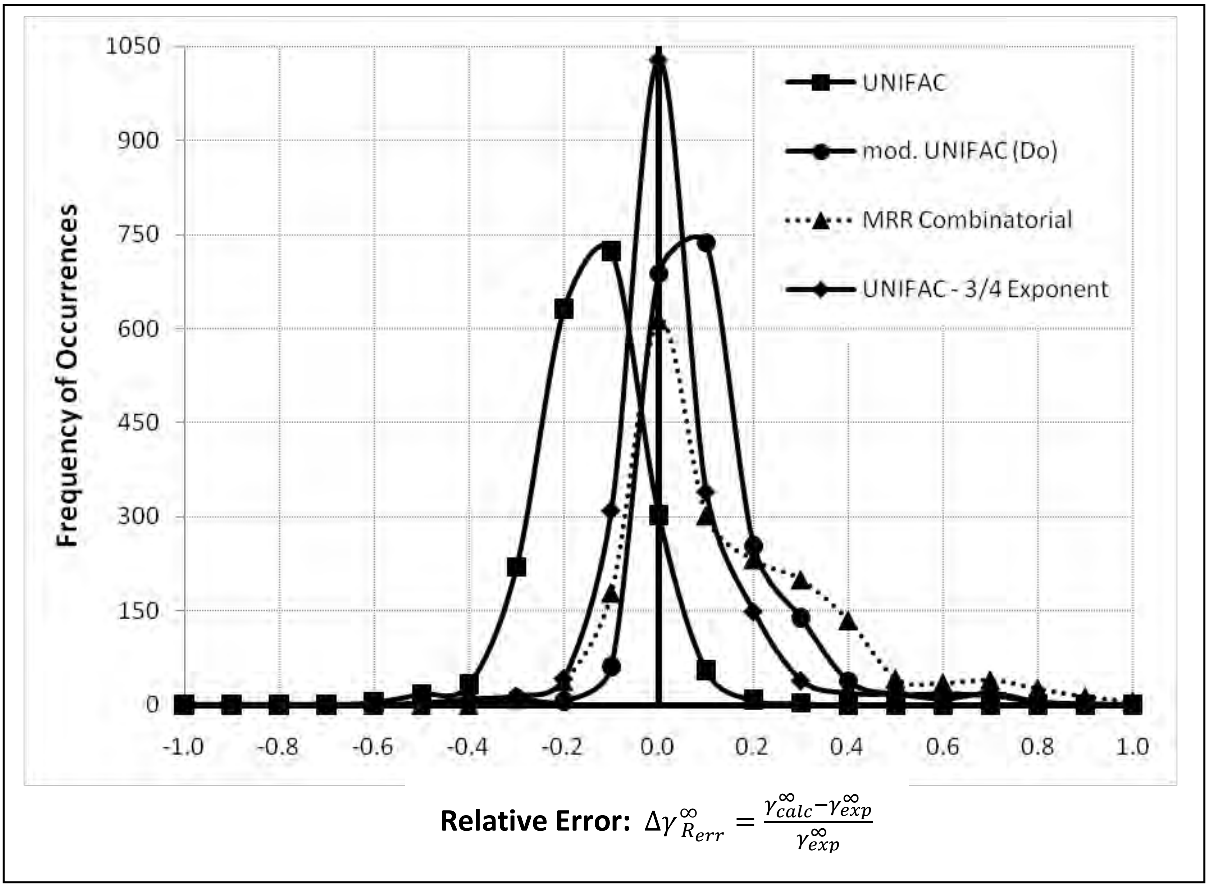


Figure 21 Distribution of relative errors in infinite activity coefficients of saturated alkane mixtures for selected GC model combinatorials, where calculated liquid activity coefficients are determined solely from combinatorial contributions (i.e. there is no contribution from the residual).

Method	Equation	Mid-point Values for the Relative Error Distributions	Percentage of Datasets Under Predicted
GS	(5-3)	-18.8 %	96.7 %
MU	(5-4)	+1.9 %	38.6 %
GS-3/4	(5-4)[9]	-4.4 %	70.3 %
MRR	(5-5)	+3.0%	42.1 %

Table 4 Mid-point values and standard deviations for the distribution curves of relative error from Figure 21 are tabulated, along with the percentage of calculated $\gamma^{\infty}_{i,calc}$ values smaller than the experimental values $\gamma^{\infty}_{i,exp}$ for each method.

The distribution of relative errors using the GS expression of the UNIFAC method resulted in a mid-point value of -18.8%, while the MU expression of mod. UNIFAC (Do.) method was calculated to have a mid-point value of 1.9% in relative error. The differences between the two distribution curves then represent the differences between the two combinatorial expressions, where the combinatorial of mod. UNIFAC (Do.) (MU, Equation (5-4)) incorporates an empirical ¾ exponent-modification and uses group volume and surface area parameters that were fitted simultaneously to

---

[9] Calculated using the group classifications and parameters defined by the UNIFAC method, which elucidates the effects of using group volume and surface area parameters as additional fitting parameters in the fitting of the mod. UNIFAC (Do.) method.

the group interaction parameters used for the model (i.e. not entirely obtained via the method of Bondi (1964) as they are in the case of the UNIFAC method).

In order to determine the effect of using the group volume and surface area parameters as fitting constants, Equation (5-4) was recalculated using the group parameters defined by the original UNIFAC method, which are determined by the method of Bondi (1964) (labelled “GS-¾” in Figure 21). The resulting differences suggests that the fitting of the group size and shape parameters seems to add only marginally to the superior fit of MU over GS, where the absolute difference of their mid-point values is 20.7% ($|1.9\% - 18.8\%|$, see Table 4). The ¾ exponent-modification, therefore, results in the largest impact on the calculated behaviour concerning the MU expression, where the absolute difference of their mid-point values is 6.3% ($|1.9\% - 4.4\%|$, see Table 4). This implies that the ¾ exponent-modification provides roughly 69.6% of the shift from the GS mid-point line to that of the MU mid-point line (where $69.6\% = (20.7\% - 6.3\%)/20.7\%$). The effect of using the group volume and surface area parameters as fitting constants, therefore, is not too significant in comparison.

The fitted size and shape factors of the MU combinatorial, however, also seem to disturb the Gaussian-like character of the distribution curve when the GS-3/4 combinatorial is compared against the MU combinatorial. This seems to support the observation (Moller, 2010) that the MU combinatorial has lost some physical significance by including the group volumes and surface area parameters in the fitting-process, and that the fluctuations in the calculated behaviour of the MU combinatorial are compensated for by the residual contribution of the mod. UNFAC (Do.) method (i.e. they are coupled).

Now comparing the distribution curve of the MRR combinatorial with the other distributions depicted in Figure 21, the curve matches reasonably well to the results of the MU and GS-¾ combinatorials. Furthermore, with a mid-point value of 2.95%, the MRR combinatorial only under-predicts the experimental data points 42.1% of the time. As expected, the MRR combinatorial is therefore suitable to alkane-solvent systems only.

5.2.1.1. *Empirical Correction*

In view of the systematic deviations exhibited by the group contribution methods, a simple correction/multiplication factor can be used to shift the distribution curves in relative error to be more or less centred on zero using the mid-points of each curve (Figure 19, tabulated in Table 3):

$$\gamma_i^{\infty,*} = \text{Factor} * \gamma_i^{\infty} \tag{5-6}$$

Where the correction factor can be defined in terms of the relative error (R_{err}) being amended:

$$\text{Factor} = \frac{1}{(1 + \text{R}_{\text{err}})} \tag{5-7}$$

Given that the application of a correction factor will simply shift the distributions, the choice of the “correct” factor is in large part dependent on the intended application; because, better results could be obtained by applying a case-specific correction (e.g. using Table 4 to correct for alkanes). Given that the correction is not the focus of the present work, proving the concept should suffice. In so doing, the applied corrections are those determined from the median point of the overall distributions (i.e. Figure 20). As shown by Figure 22 these corrections improve the ability of these methods to predict infinite activity coefficient data, and is quite simple to implement (i.e. no system-specific information is required in order to apply the correction).

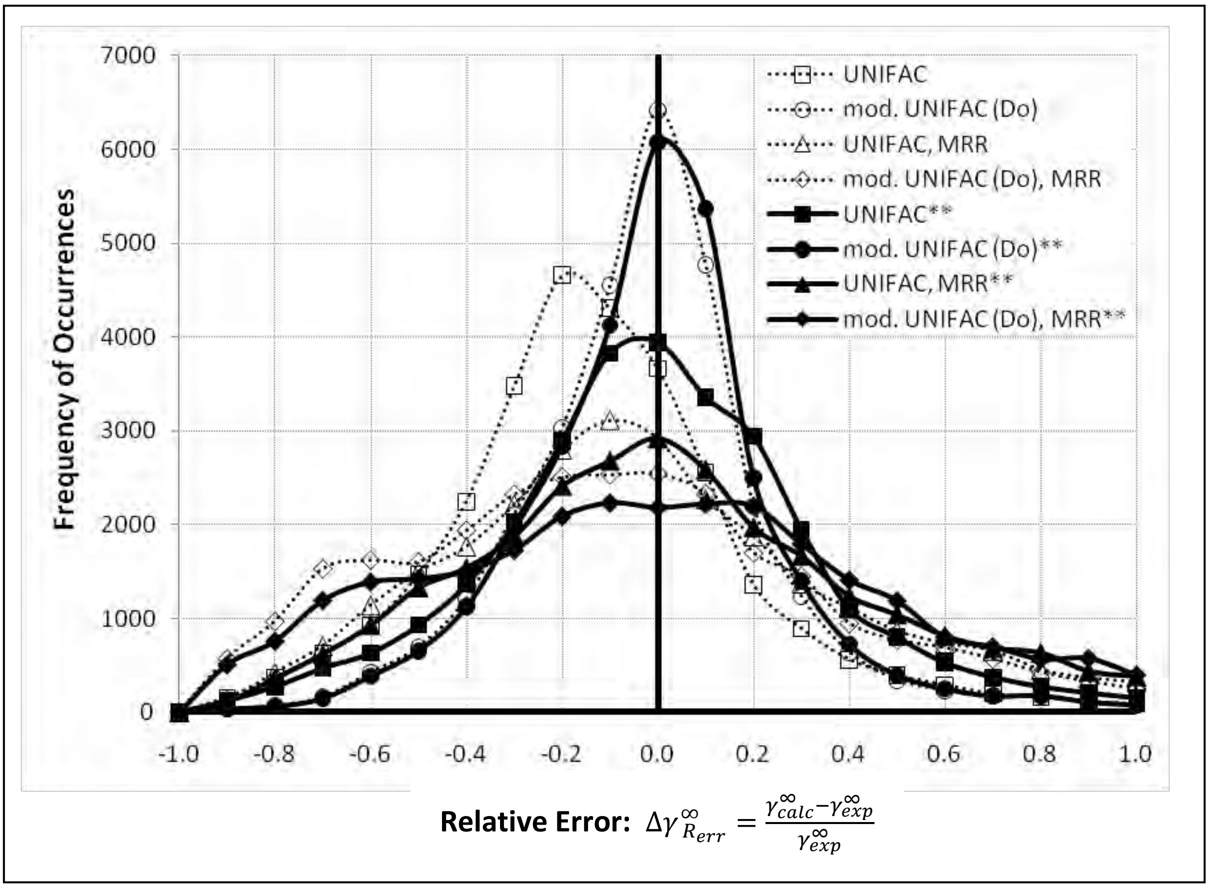
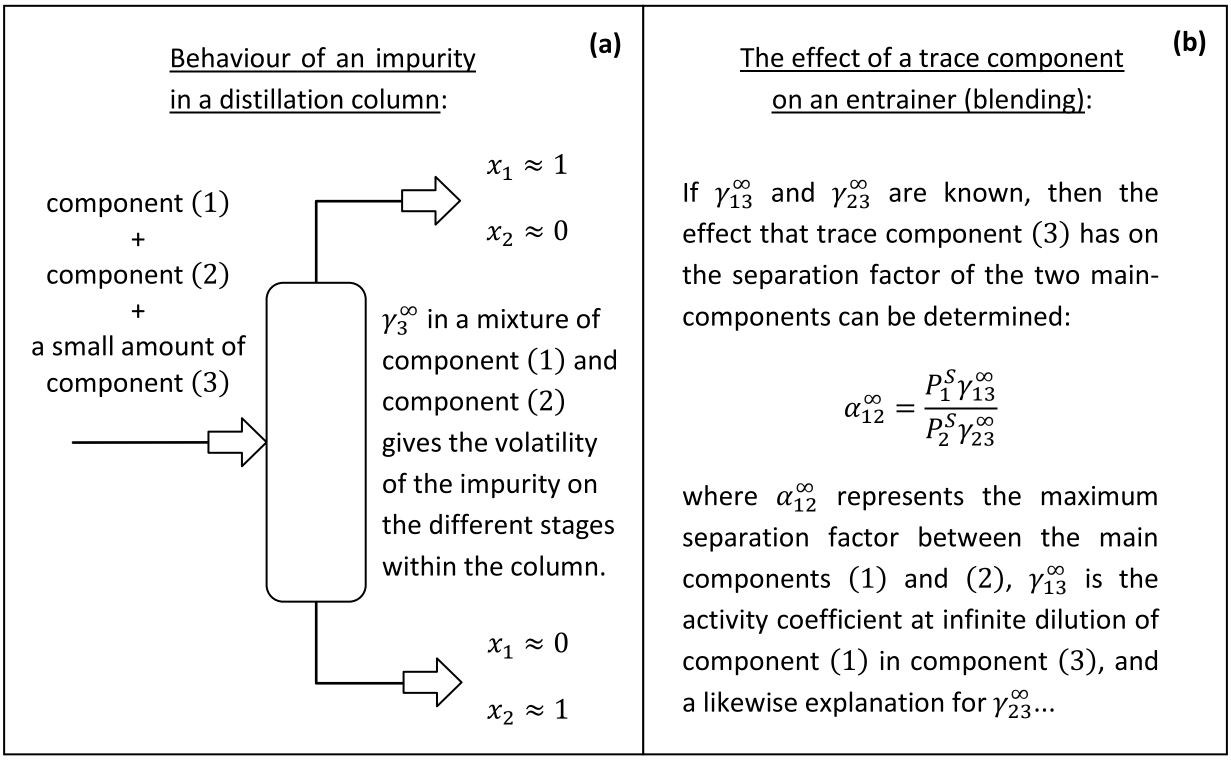


Figure 22 Distribution curves of relative errors in activity coefficients at infinite dilution with applied correction factors (Equation (5-6) and Equation (5-7)) obtained from predicting DDB-ACT data using UNIFAC and mod. UNIFAC (Do.) methods (2008 Consortium parameters), including the variations obtained by directly replacing the combinatorial expressions in each equation with the MRR combinatorial.

5.3. Ability to Represent Multicomponent Mixtures

Practical applications like the modelling and/or design of distillation columns in most cases require sufficiently precise descriptions of the behaviour of mixtures containing more than two components. These descriptions are usually calculated using g^E-models (e.g. Wilson, UNIFAC, etc.) or equations of state, and typically only require pure component and binary parameters. The other higher-order interactions are generally neglected due to the scarcity and difficultly of measuring good ternary and higher data necessary for the fitting of any higher-order interactions between mixture molecules (as mentioned in Chapter 3.1). This means, for instance, that it is rarely possible to test the accuracy of ternary (and higher) predictions against actual experimental data containing more than two components, although the true nature of multicomponent mixtures is of industrial importance.

One of the few sources of experimental information available to evaluate this effect is the ACM databank of DDBST, where mixture composition effects on the behaviour of a trace component can be evaluated (DDBST Software and Separation Technology GmbH, 2010). This type of behaviour can be very important for the design and operation of many chemical processes. In the case of manufacturing reaction grade chemicals, for instance, the concentrations of specific impurities could mean the difference between meeting and missing targeted production purities. In other cases impurities may cause operational problems, where designed rates and separations cannot be reached. An illustration of the behaviour of an impurity in a distillation column is presented in the left-hand side of Figure 23 below.



If γ_{13}^∞ and γ_{23}^∞ are known, then the effect that trace component (3) has on the separation factor of the two main-components can be determined:

$$\alpha_{12}^\infty = \frac{P_1^S \gamma_{13}^\infty}{P_2^S \gamma_{23}^\infty}$$

where α_{12}^∞ represents the maximum separation factor between the main components (1) and (2), γ_{13}^∞ is the activity coefficient at infinite dilution of component (1) in component (3), and a likewise explanation for γ_{23}^∞...

Figure 23 Illustrations showing applications of why DDB-ACM data, and the calculation of such values, is important in process engineering. Figure-left shows how such data and its calculation is important for the description of an impurity in a distillation column, and figure-right shows the effect that a trace component has on the blending of an entrainer (where the trace component influences the maximum separation factor between component (1) and component (2)).

This type of information is also used to evaluate the effectiveness of entrainer blends, where small amounts of a component can potentially affect the selectivity of a known entrainer. The right-hand side of Figure 23 shows how this information is used to determine the effect that a trace component may have on the separation factor between a mixture of component (1) and component (2). If it was desired to separate an aromatic compound such as benzene (1) from n-methyl-2-pyrrolidone (NMP) (2), for instance, then water (3) could be used to influence the separation. Since water is a highly polar molecule, it can be used to influence polarity of the solution, which would then in turn affect the capacity of NMP to hold onto the non-polar benzene molecules. In other words the separation factor between benzene (1) and NMP (2) would increase, due to the increase of the activity coefficient at infinite dilution of benzene (1) in water (3) relative to NMP in water (see the right-hand side of Figure 23).

In cases like these, it becomes important to validate the ability of thermodynamic models to correctly describe these effects. Given the relevance to industry, the predictive methods were tested for their ability to represent/predict the influence that an additional component would have on a binary mixture to be separated; to this end, the following constraints were used to ensure that datasets of sufficient quality were used for the evaluation:

The constraints of Chapter 5.2 (Ability to Predict Activity Coefficients at Infinite Dilution) were also used here; except for the last constraint, requiring γ^{∞} values ≤ 500, since this constraint would further restrict an already limited database (in comparison).

Only datasets which include both end-points were used, so that the calculated/predicted solute activity coefficients at infinite dilution in the ternary mixture could be decoupled from the binary results. By shifting the calculated values to match the measured solute activity coefficients at infinite dilution in pure solvent-1 and pure solvent-2 (γ_{31}^{∞} and γ_{32}^{∞} respectively), the behaviour of the solute *in the mixture* can be elucidated.

To ensure that the general trend/behaviour of each dataset could be determined, only datasets having five or more data points were included for evaluation purposes.

The restrictions resulted in a final database consisting of 517 datasets (4,795 experimental γ^{∞} values) from 278 unique ternary-systems (of the original 1,409 datasets, from 739 unique ternary-systems, within the ACM database); a breakdown of the data available for analysis according to measurement type and measurement method is show in Table 5 below:

	ACM Datasets	ACM Datasets Available for Analysis
Measurement Method		
Dilutor	251 (1,042)	11 (74)
Ebulliometry	1 (7)	0 (0)
GLC (not specified)	22 (113)	21 (108)
GLC (with gas phase corr.)	342 (1,898)	12 (228)
GLC (without gas phase corr.)	727 (5,566)	459 (4,310)
LLC	16 (96)	--------
Other Techniques	6 (15)	0 (0)
Static Method	44 (472)	14 (75)
Measurement Types		
ISO = 0 (dynamic)	76 (304)	--------
ISO = 1 (constant x_i)	203 (889)	--------
ISO = 2 (constant T)	1130 (8,016)	517 (4,795)
TOTAL		
	1409 (9,209)	517 (4,795)
Parameter Availability		
UNIFAC		266 (2,298)
mod. UNIFAC (Do.)		289 (2,412)
Removed/Filtered Data		
Water + hydrocarbon systems	216 (982)	
Number of data points < 5	460 (1,576)	
γ_{31}^{∞} and γ_{32}^{∞} not measured	760 (3,918)	

Table 5 Breakdown of the datasets and data points (in parentheses) of the ACM database of DDB and the subset of data used for analysis, according to measurement type and measurement method.

5.3.1. Calculations

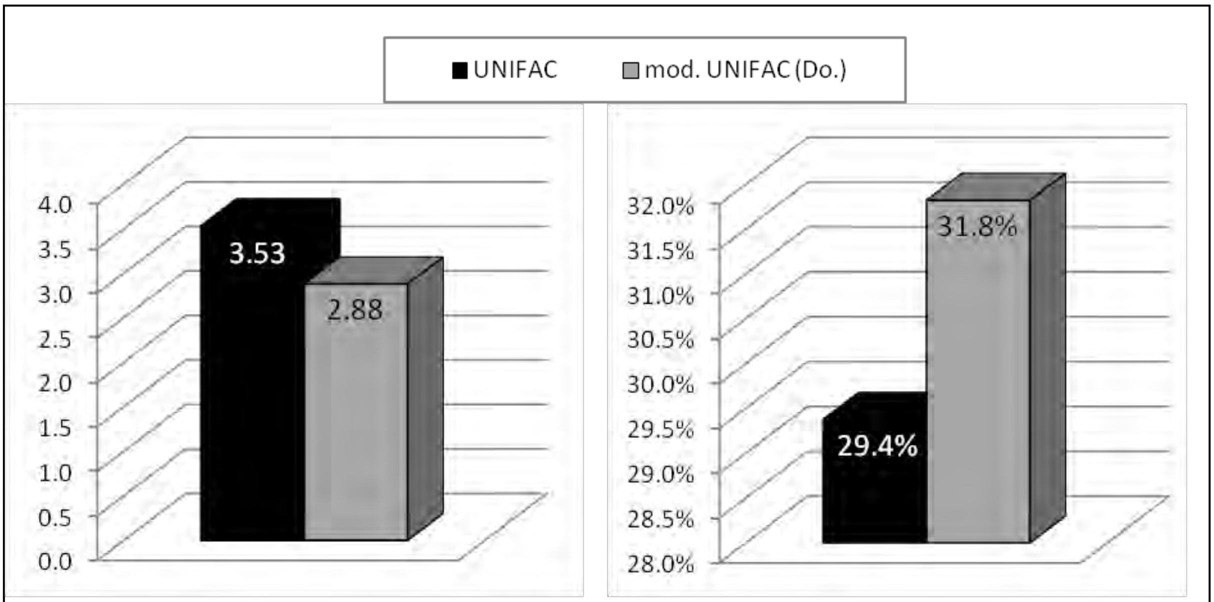
Accurate representation of mixtures is required for the design, operation, and evaluation of many plant operations (such as column stripping sections, where the focus is on removing trace impurities). Furthermore, practical problems are rarely limited to binary systems, so the ability of group methods to accurately predict this information can be considered industrially important. Given that no broad study was found in literature, a comparison of the efficiency of these methods to represent DDB-ACM data stored in the Dortmund Data Bank (DDBST Software and Separation Technology GmbH, 2009) was performed.

The activity coefficients at infinite dilution of solutes in mixtures were predicted with the group contribution equations of UNIFAC and mod. UNIFAC (Do.) only, since the MRR combinatorial was only developed and is applicable for describing binary mixtures. For comparison, the relative errors were calculated using Equation (5-1) for each data point. Furthermore, the relative-average-deviations (RAD) of the calculated results, with respect to the experimental data, were calculated using the following definition:

$$\Delta\left(\gamma_{3,solvent}^{\infty}\right)_{RAD} = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\left(\left(\gamma_{i,solvent}^{\infty}\right)_{calc} - \left(\gamma_{i,solvent}^{\infty}\right)_{exp}\right)}{\left(\gamma_{i,solvent}^{\infty}\right)_{exp}}\right] \tag{5-8}$$

where n is the number of data points and $\gamma_{i,calc}^{\infty}$ and $\gamma_{i,exp}^{\infty}$ are the calculated and experimental activity coefficients at infinite dilution. The overall RAD value of each method was determined without the inclusion of the end-points (i.e. data points representing the solute in pure solvent-1 or pure solvent-2); since the inclusion of these values would mask the ability of the group contribution models to correctly predict the real behaviour of a dilute-solute in a "mixture".

The resulting RAD values are depicted graphically in Figure 24, along with the averaged absolute deviations in activity coefficients at infinite dilution. The figure shows that the mod. UNIFAC (Do.) method results in the best description of the real behaviour of a very dilute component (solute) in a mixture, followed reasonably close by the results of the UNIFAC method. The differences that exist between the results of UNIFAC and mod. UNIFAC (Do.) in Figure 24, then represent the differences that exist between the two methods. Differences in the combinatorial expressions, the way in which each method component fragmentations, and the parameters used in each of the methods all contribute to the differences in the results depicted in Figure 24.



Absolute Deviation:

$$\frac{1}{n}\left[\sum_{i=1}^{n}\left|\left(\gamma_{3,solvent}^{\infty}\right)_{calc}-\left(\gamma_{3,solvent}^{\infty}\right)_{exp}\right|\right]$$

Relative Average Deviation (RAD):

$$\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\left(\left(\gamma_{i,solvent}^{\infty}\right)_{calc}-\left(\gamma_{i,solvent}^{\infty}\right)_{exp}\right)}{\left(\gamma_{i,solvent}^{\infty}\right)_{exp}}\right]$$

Figure 24 Overall absolute deviations and relative-average deviations (RAD) of infinite dilution activities over the filtered DDB-ACM, where the relative errors of the solute in pure solvent-1 or pure solvent-2 (γ_{31}^{∞} and γ_{32}^{∞}) have been omitted.

In an effort to understand what influences these differences have on the calculated results, distribution curves of the calculated relative error (see Equation (5-1)) were made for UNIFAC and mod. UNFAC (Do.), where the calculated activity coefficients at infinite dilution (γ_{calc}^{∞}) were determined strictly from the residual contributions of each method (i.e. determined solely from the solution of groups concept). The distribution curves determined solely from residual contributions are shown in Figure 25, while the standard distribution curves (combinatorial and residual contributions) of UNIFAC and mod. UNIFAC (Do.) are shown in Figure 26. The differences between these two curves can then provide an indication of the how the combinatorial expressions affect the calculated results, by comparing the two curve-types (i.e. the residual contributions of Figure 25 versus the overall distributions of Figure 26).

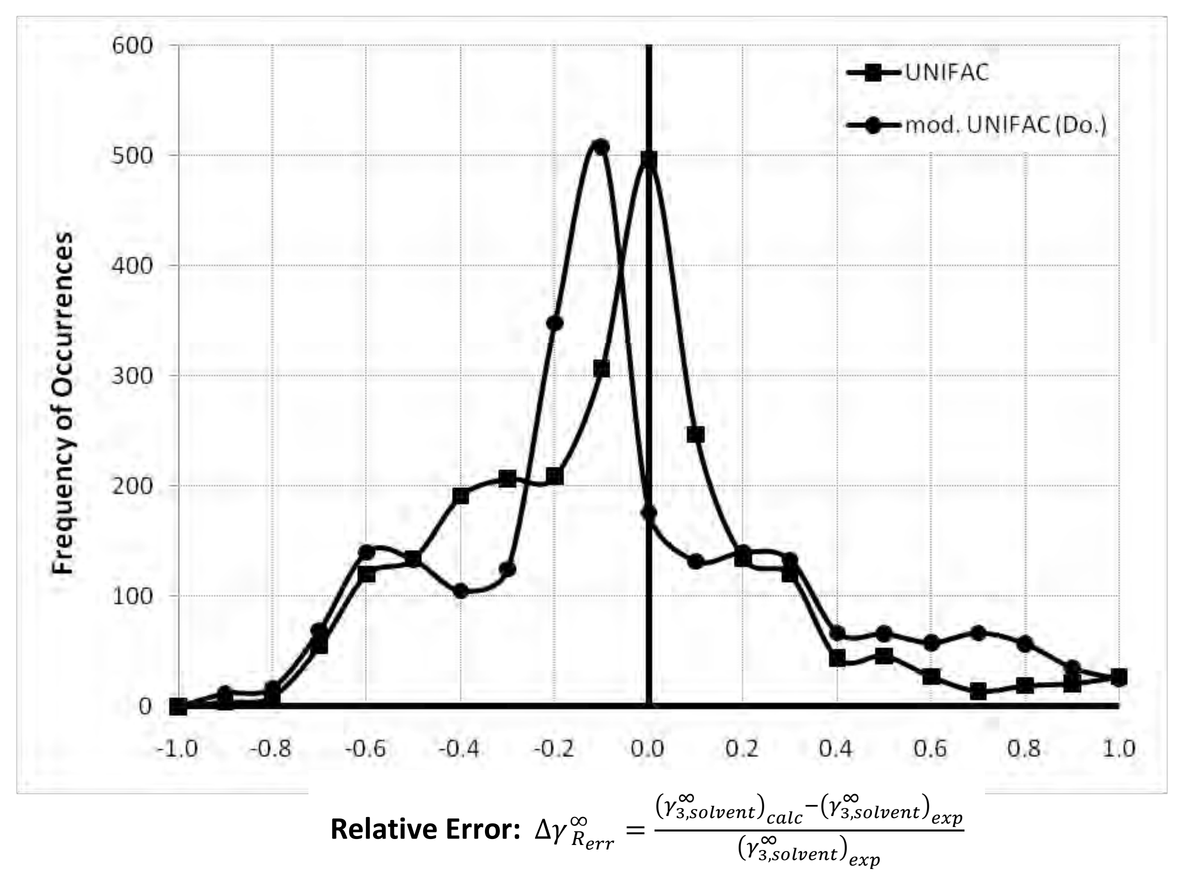
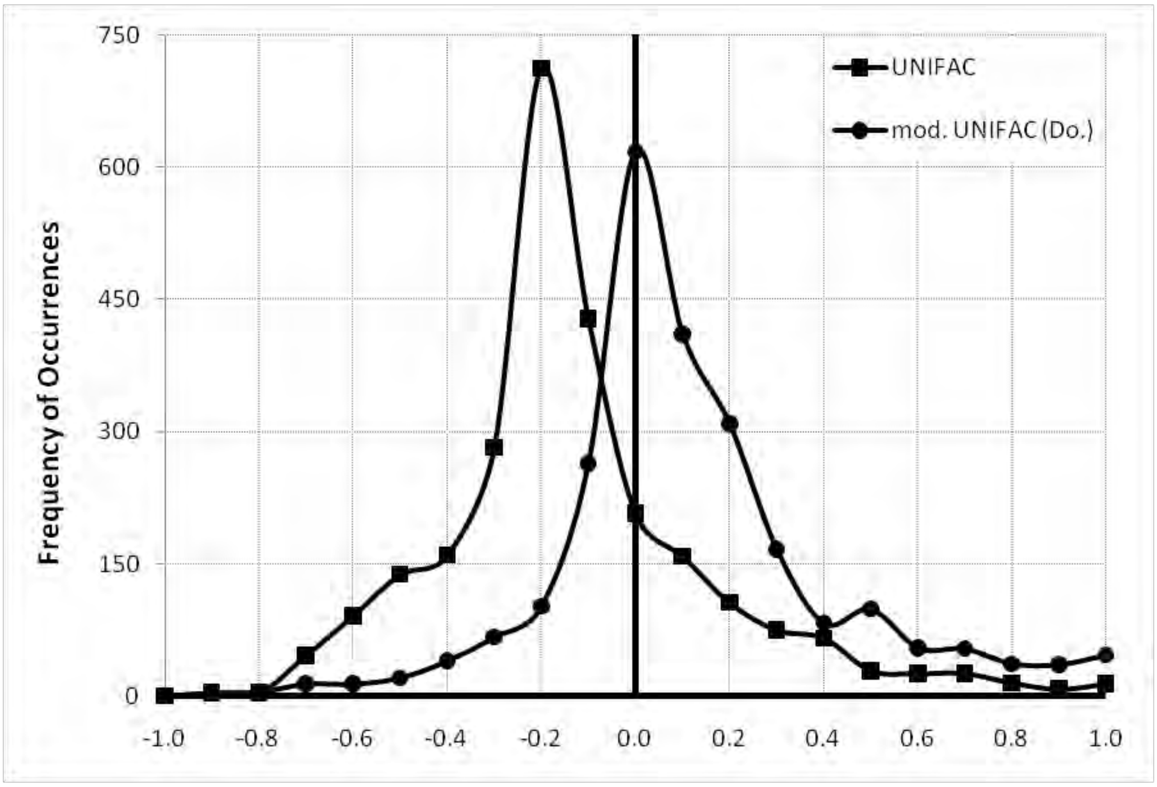


Figure 25 Distribution of relative errors of solute activity coefficients in mixtures for selected group contribution methods (using only residual contributions), where the relative errors of the solute in pure solvent-1 or pure solvent-2 (γ^{∞}_{31} and γ^{∞}_{32}) have been omitted.



$$\textbf{Relative Error: } \Delta\gamma^{\infty}_{R_{err}} = \frac{(\gamma^{\infty}_{3,solvent})_{calc}-(\gamma^{\infty}_{3,solvent})_{exp}}{(\gamma^{\infty}_{3,solvent})_{exp}}$$

Figure 26 Distribution of relative errors of solute infinite activity coefficients in mixtures for selected group contribution methods, where the relative errors of the solute in pure solvent-1 or pure solvent-2 (γ^{∞}_{31} and γ^{∞}_{32}) have been omitted.

It should be noted that the curves of Figure 25 and Figure 26 do not include the relative errors resulting from the solute in pure solvent-1 or the solute in pure solvent-2 (γ^{∞}_{31} and γ^{∞}_{32} respectively), since the inclusion of these values would mask the ability of these methods to predict the solute activity coefficients at infinite dilution in the "mixture." Furthermore, the mid-point values of each distribution curve in Figure 25 and Figure 26 have been tabulated in Table 6 for convenience.

Method	Figure	Mid-point Values for the Relative Error Distributions	Percentage of Data Points Under Predicted
UNIFAC	Figure 26	-21.8 %	78.4 %
UNIFAC - *Residual*	Figure 25	-10.5%	65.6 %
mod. UNIFAC (Do.)	Figure 26	1.7 %	43.5 %
mod. UNIFAC (Do.) - *Residual*	Figure 25	-16.0 %	61.8 %

Table 6 Mid-point values for the distribution curves of relative error from Figure 25 and Figure 26 are tabulated, along with the percentage of calculated $(\gamma^{\infty}_{i,solvent})_{calc}$ values smaller than the experimental values $(\gamma^{\infty}_{i,solvent})_{exp}$ for each method.

In comparing the distributions and mid-point values of Figure 25, we see that UNIFAC represents the DDB-ACM data slightly more accurately than mod. UNIFAC (Do.), where both distribution curves were solely calculated based on the contributions from the solution of groups of each method (i.e. the residual terms). The differences, therefore, represent the differences in the parameters used to represent the characteristic interactions between group pairs, where the mod. UNIFAC (Do.) method uses temperature-dependent parameters. Since the data used for the distributions were distributed fairly evenly across a temperature range of 275 – 350 K, the lower mid-point value from the residual of mod. UNIFAC (Do.) likely arises from the use of multiple data-types in the fitting procedure (i.e. an averaging of errors across additional data compared to UNIFAC).

In comparing the distributions of Figure 25 and Figure 26, we can see the effect that the combinatorial contribution has on the residual terms. Given the good distribution shown by the distribution of UNIFAC in Figure 25, it is not too surprising that poor distributions are obtained for UNIFAC as shown in Figure 26. Any combinatorial contribution other than unity would likely result in a worsening effect for UNIFAC. As for the mod. UNIFAC (Do.) method (with a mid-point value of 1.7%, as shown in Figure 26), the combinatorial expression correctly adjusts for the residual contribution of the method. The combinatorial contribution of mod. UNIFAC (Do.), therefore, has the largest impact on the final calculated results using the method (shifting the mid-point by 17.8% versus the 11.3% shift in the case of UNIFAC).

In order to elucidate the qualitative ability of the predictive methods to represent the real phase behaviour of the DDB-ACM data, the results were decoupled from binary influences by shifting the calculated results to match the experimental activity coefficients at infinite dilution of the solute in pure solvent-1 (γ_{31}^{∞}) and of the solute in pure solvent-2 (γ_{32}^{∞}). Figure 27 and Figure 28 are used to illustrate this using the system of 2-methyl-2-butene (3) in a mixture of 1-methyl naphthalene (1) and glutaronitrile (2), where the following steps have been employed:

1. Values for both the experimental and calculated solute activity coefficients at infinite dilution in the mixture are determined from Equation (5-9) and Equation (5-10), where the solute activity coefficient at infinite dilution in the mixture ($\gamma_{3,solvent}^{\infty}$) is assumed to be a mole-fraction-weighted sum of the solute in each of the pure solvents at infinite dilution (γ_{31}^{∞} and γ_{32}^{∞}). The resulting ideal solution lines (γ_{exp}^{*} and γ_{calc}^{*} respectively) are depicted in Figure 27 as dashed lines between the pseudo-experimental and pseudo-calculated end points.
2. The calculated curvature (or deviation) is determined using Equation (5-11), which represents the "shift" from the calculated ideal solution line determined via Equation (5-10). The calculated shift ($\Delta\gamma_{calc}$) is also depicted in Figure 27 for illustration purposes.
3. The final "shifted" values $\left(\gamma_{3,solvent}^{\infty}\right)_{shift}$ are then determined via Equation (5-12) for each calculation method being investigated; with respect to the illustration of Figure 27, the calculation results have been shifted as shown in Figure 28.

As Figure 28 shows, the shifted results for the mod. UNIFAC (Do.) method retains the character of the original predictions (as illustrated in Figure 27), but have been decoupled from the binary results by shifting the calculated end points to match the experimental values.

$$\gamma^*_{exp} = x_1\gamma^{\infty}_{31,exp} + x_2\gamma^{\infty}_{32,exp} \tag{5-9}$$

$$\gamma^*_{calc} = x_1\gamma^{\infty,}_{31,calc} + x_2\gamma^{\infty}_{32,calc} \tag{5-10}$$

$$\Delta\gamma_{calc} = \left[\left(\gamma^{\infty}_{3,solvent}\right)_{calc} - \gamma^*_{calc}\right] \tag{5-11}$$

$$\left(\gamma^{\infty}_{3,solvent}\right)_{shift} = \left(\gamma^{\infty}_{3,solvent}\right)_{exp} + \Delta\gamma_{calc} \tag{5-12}$$

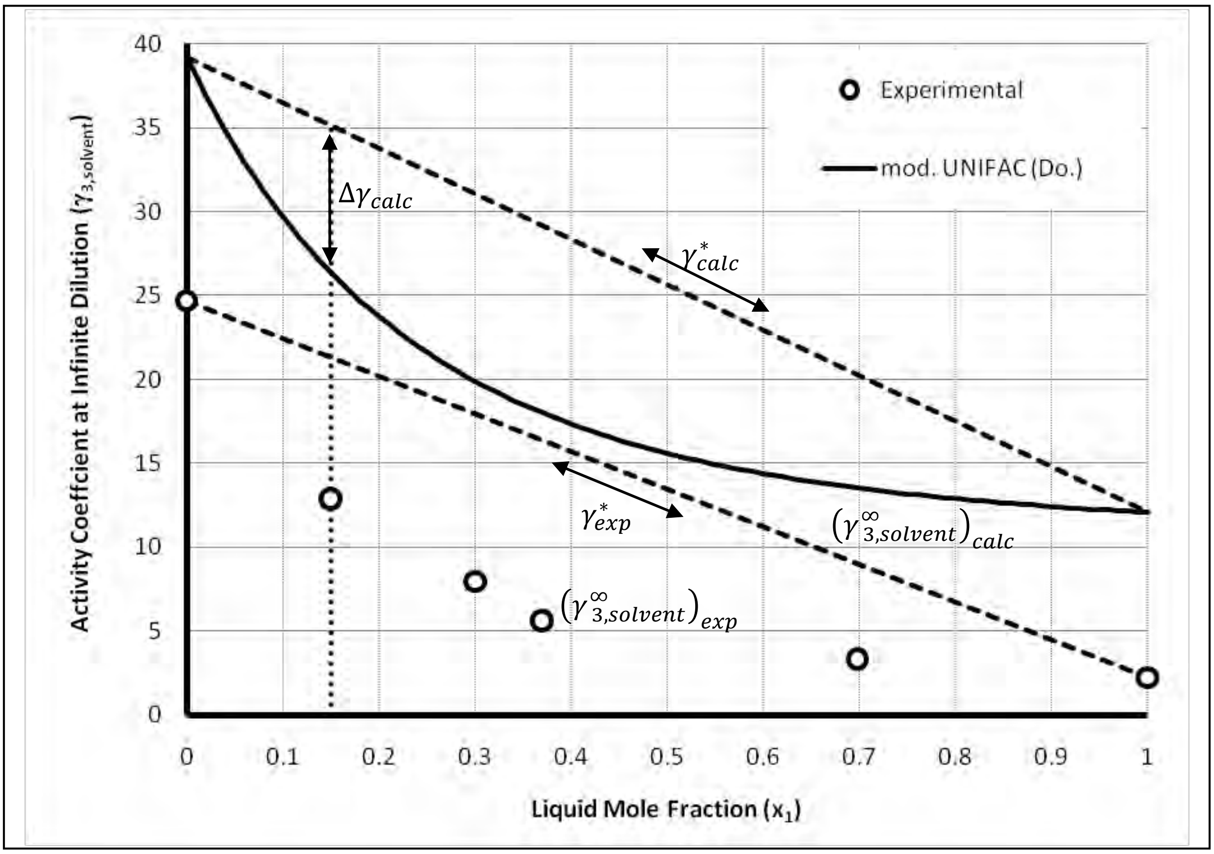


Figure 27 DDB-ACM dataset number 18, consisting of 1-methyl naphthalene (1) + glutaronitrile (2) + 2-methyl-2-butene (3) at 298.15 K. This figure shows the values of γ^*_{exp} calculated using Equation (5-9), γ^*_{calc} values determined using Equation (5-10), and the $\Delta\gamma_{calc}$ values obtained using Equation (5-11).

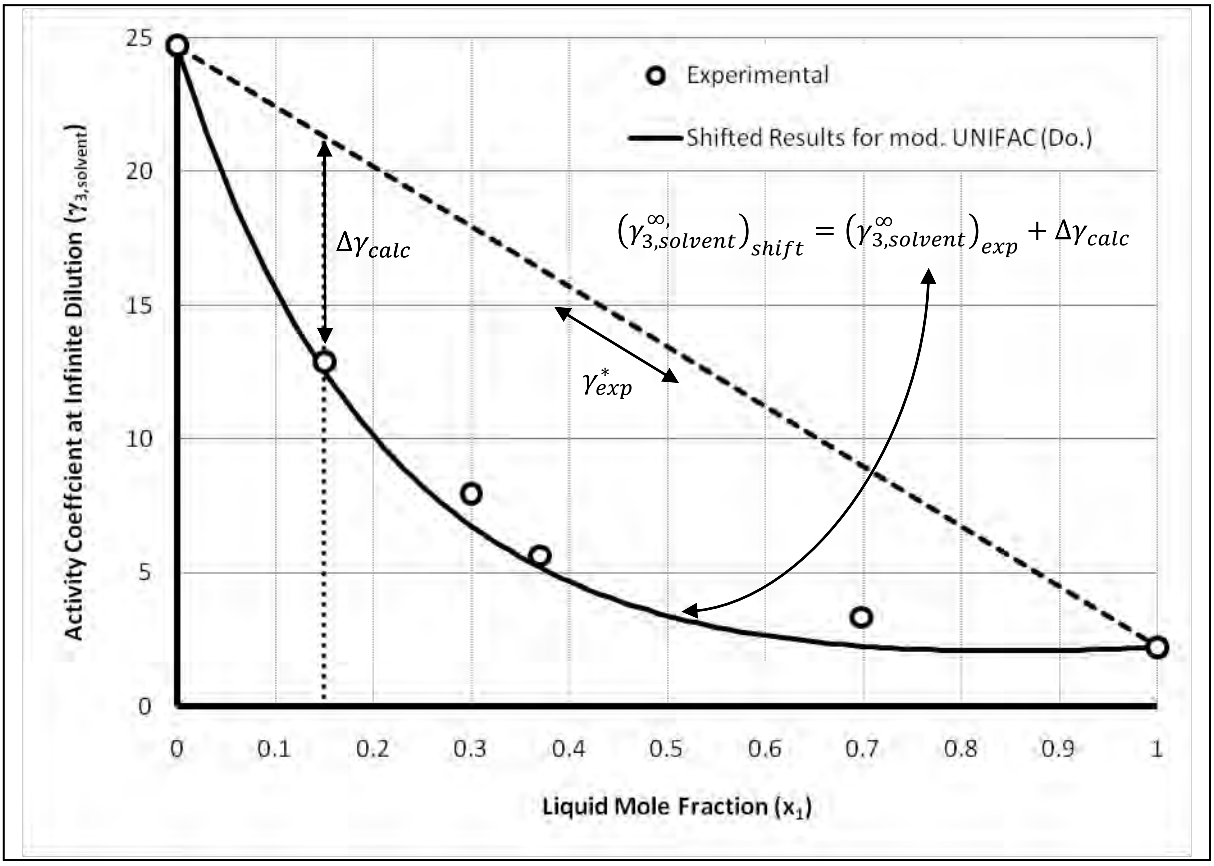


Figure 28 DDB-ACM dataset number 18, consisting of 2-methyl-2-butene (3) in a mixture of 1-methyl naphthalene (1) and glutaronitrile (2) at 298.15 K. This figure shows the values of γ^{*}_{exp} calculated using Equation (5-9), $\Delta\gamma_{calc}$ values determined using Equation (5-11), and the desired shifted model results $\left(\gamma^{\infty}_{3,\text{solvent}}\right)_{\text{shift}}$ obtained using Equation (5-12).

The resulting distributions in relative errors (shifted) are shown in Figure 29 for each group contribution method. The relative errors corresponding to the solute activity coefficients at infinite dilution in pure solvent-1 and pure solvent-2 (γ^{∞}_{31} and γ^{∞}_{32} respectively) were not included, since the shifted relative errors at these points have a value of zero, and would therefore greatly skew the distributions. When looking at Table 7, the UNIFAC method is found to have the closest mid-point value to zero (with a value of -0.3%) followed closely by the mid-point value of mod. UNIFAC (Do.) (with a value of of-0.6%). When Figure 29 is considered, however, it is obvious that the marginally better mid-point value of -0.3% for UNIFAC is due to favourable averaging. It is clear from the distributions that the mod. UNIFAC (Do.) method provides the best overall description of the shifted-values (see Figure 29); both methods however, seem to reproduce the characteristic-shape of the data reasonably well once the calculated end-points are shifted to match the experimental end-points.

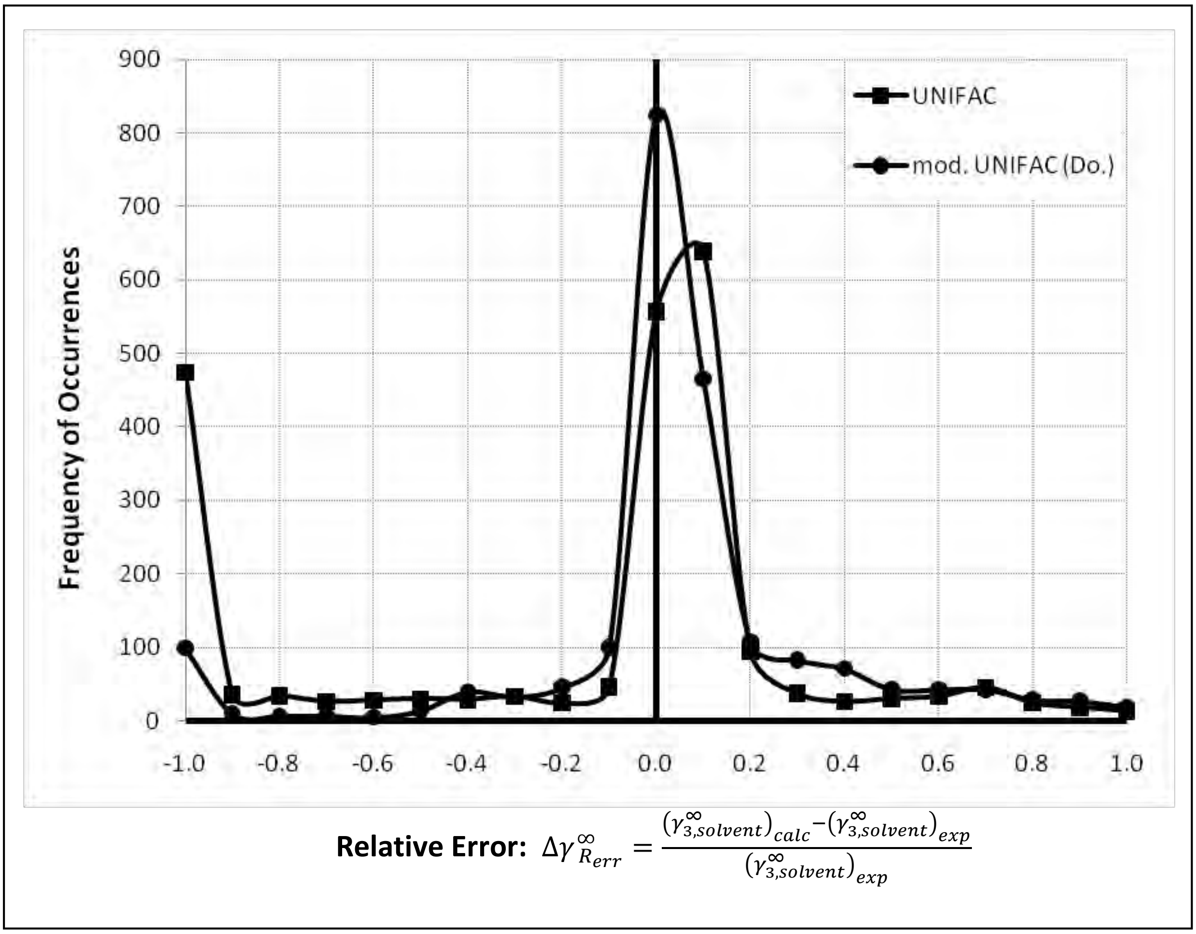


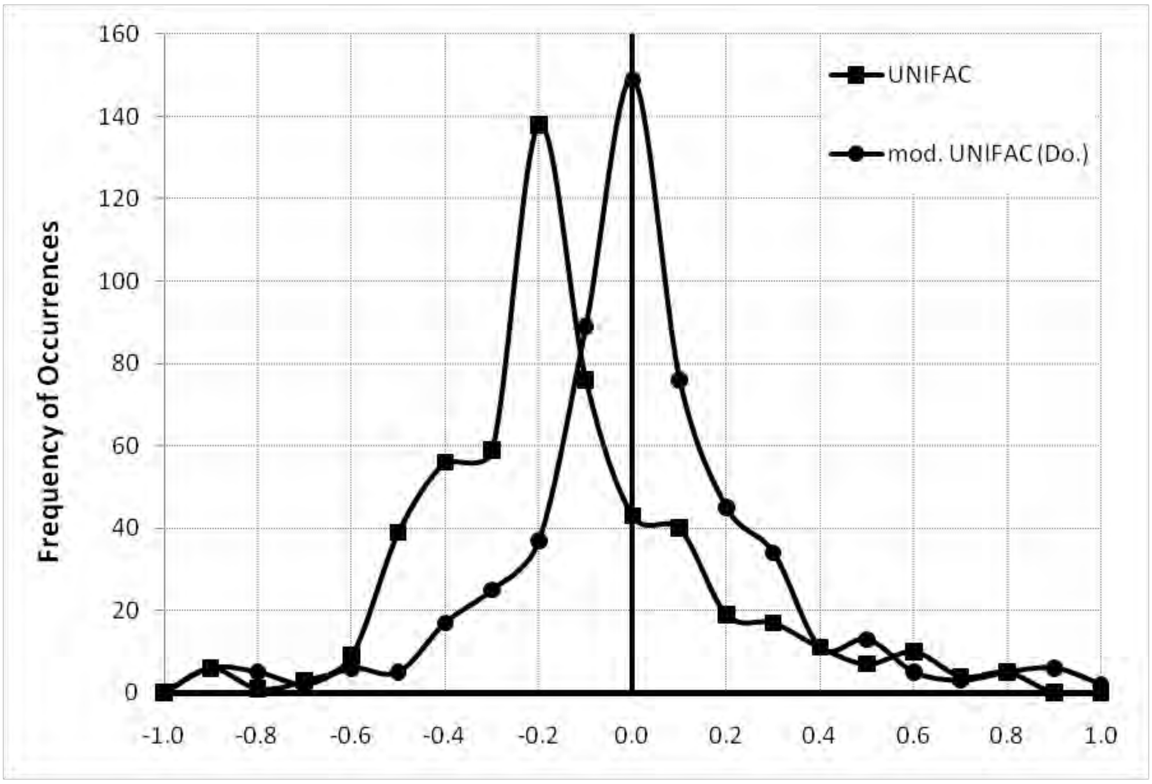
Figure 29 Distribution of relative errors (shifted) of solute infinite activity coefficients in mixtures for selected group contribution methods, where the relative errors of the solute in pure solvent-1 or pure solvent-2 (γ^{∞}_{31} and γ^{∞}_{32}) were not included.

Method	Mid-point Values for the Relative Error Distributions	Percentage of Datasets Under Predicted
UNIFAC	-0.3 %	49.9 %
mod. UNIFAC (Do.)	-0.6 %	44.9 %

Table 7 Mid-point values for the distribution curves of relative error from Figure 29 are tabulated, along with the percentage of calculated $\gamma^{\infty}_{3,calc}$ values smaller than the experimental values $\gamma^{\infty}_{3,exp}$ for each method.

The fact that there appears to be a large amount of shifted data points resulting in large negative deviations in Figure 29 (those with shifted-values having $\mathrm{R_{err}} \leq -90\%$), they are found to be made mostly of solute systems containing alcohols, where one of the solvents is either another alcohol (including methanol), an ether, or a cyanide. These combinations cause over predictions for the activity coefficient at infinite dilution for some of these alcohol-solvent pairs (γ^{∞}_{31} and/or γ^{∞}_{21}), which cause large negative shifts to take place when the results are decoupled from the binary-influences. As can be seen by Figure 30 and Table 8, however, the over prediction of the activity coefficients at infinite dilution for the solute in pure solvent-1 and pure solvent-2 are rather localized. In fact, when the gross negative values in Figure 29 are analyzed further, the values are accounted for by only 11% of the datasets for UNIFAC and only 6% of the datasets for mod. UNIFAC (Do.). The larger percentage of datasets resulting in negative deviations for the UNIFAC method was found to be caused by the over predictions of γ^{∞} values for alcohols in pure methanol (compared to mod. UNIFAC (Do.)). An example of this is shown in Figure 31 for the system methanol (1) + water (2) + ethanol (3), where the shifted UNIFAC values actually result in negative activity coefficients. Figure 31 illustrates that the degree and the direction of the shift depends on how accurate the experimental activity coefficients at infinite dilution of the solute in pure solvent-1 (γ_{31}^{∞}) and of the solute in pure solvent-2 (γ_{32}^{∞}) are predicted. Therefore, similar negative shifts will likely result anytime these endpoints are greatly over predicted or large positive shifts when these points are greatly under predicted.

Overall, however, distributions of Figure 29 and Figure 30 show the mod. UNIFAC (Do.) method represents the DDB-ACM database with more accuracy and precision than the original UNIFAC method. These results give further justification for using the mod. UNIFAC (Do.) method as the base-GC method in the proposed hybrid method, and contributes new knowledge of the ability of these methods to represent multicomponent mixtures. Furthermore, as additional experimental information is added to the DDB-ACM database, relationships can perhaps be identified for specific mixture-types; at present, however, no such relationships have been identified.



$$\textbf{Relative Error: } \Delta\gamma^{\infty}_{R_{err}} = \frac{(\gamma^{\infty}_{3,solvent})_{calc} - (\gamma^{\infty}_{3,solvent})_{exp}}{(\gamma^{\infty}_{3,solvent})_{exp}}$$

Figure 30 Distribution of relative errors of solute infinite activity coefficients in pure solvent-1 (γ^{∞}_{31}) and pure solvent-2 (γ^{∞}_{32}) for selected GC methods.

Method	Mid-point Values for the Relative Error Distributions	Percentage of Datasets Under Predicted
UNIFAC	-22.2 %	76.5 %
mod. UNIFAC (Do.)	-4.7 %	60.7 %

Table 8 Mid-point values for the distribution curves of relative error from Figure 30 are tabulated, along with the percentage of calculated $\gamma^{\infty}_{3,calc}$ values smaller than the experimental values $\gamma^{\infty}_{3,exp}$ for each method.

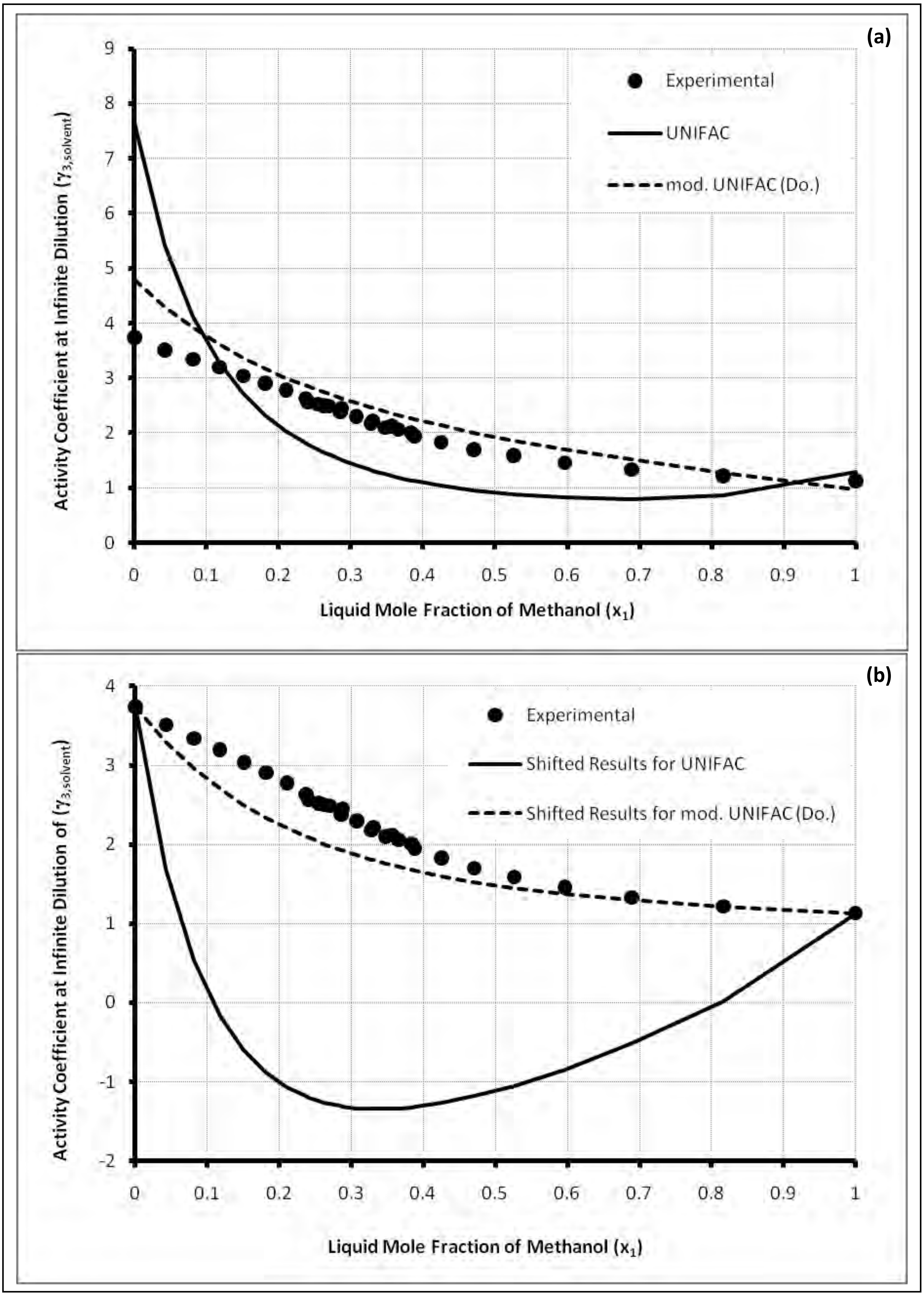


Figure 31 DDB-ACM dataset number 710, consisting of ethanol (3) in a mixture of methanol (1) + water (2) at 298.15 K. The top-figure is the non-shifted predictions using the UNIFAC and mod. UNIFAC (Do.) methods, while the shifted-values for both methods are shown in the bottom-figure.

5.4. GC2gE Conversion: *Loss in Translation*

As mentioned, models like Wilson, NRTL and UNIQUAC have long been used with great success, but their applicability is restricted by the limited availability of binary interaction parameters. Out of this reason, group contribution methods like ASOG, UNIFAC, and mod. UNIFAC (Do.) have gained wide-spread use in industry. If VLE data are not available, binary parameters can be estimated based on the results of these predictive methods. Group interaction parameters are regressed, however, to reproduce the behaviour of a large number of different mixtures simultaneously—often more than 3000 experimental data points are used to obtain the parameters for one group pair (Gmehling, 1995)—they are rarely capable of representing specific experimental information more accurately than component-based models fitted directly to individual datasets. Predictive methods, therefore, are considered second-choice data generators, where the predicted data are regressed in the same way as true experimental results in order to obtain the required component-based interaction parameters used by component-based equations like NRTL.

Predictive methods are often used in this way as "additional data sources," and very seldom are these methods directly employed during process simulation. In doing so, however, one may introduce unnecessary errors when translating the predictive results into one of the component-based models. This Chapter attempts to quantify this by comparing the predictions of group contribution models against component-based models fitted to these same predictions. The number of binary systems in this evaluation is only limited by the number of binary interaction parameters between maingroups, and by the availability of subgroups to correctly represent the system of components for the systems being considered (i.e. if the system can be predicted, parameters can be fitted to the predictions).

In order to restrict the test to a meaningful number of binary mixtures, only DDB-VLE systems matching the criteria given in Chapter 5.1 were used. The following additional criteria were also imposed so that the results could be compared to experimental datasets—binary Px(T)—of sufficient quality:

Data for water-hydrocarbon mixtures were not considered due to well-known shortcomings of the mod. UNIFAC (Do.) method.

Data for water-amine mixtures were also not considered since the fitting results for these systems resulted in very large deviations, which greatly skewed the overall statistics.

Data for systems containing carboxylic acids and hydrogen fluoride (HF) were removed so that consistent ideal vapour phase descriptions could be obtained (carboxylic acids and HF generally require using a chemical theory in order to obtain correct vapour phase descriptions).

The mod. UNIFAC (Do.) model was used as the group contribution model of choice since the method gives the best overall performance when compared against available experimental data (recall the γ^{∞} evaluation using DDB-ACT in Chapter 5.2, and the assessment of DDB-ACM presented in Chapter 5.3). Pseudo experimental datasets were then predicted using the mod. UNIFAC (Do.) method for each DDB-VLE dataset satisfying the criteria above. These predicted datasets were generated at four different temperatures over the entire composition range, in an attempt to capture the temperature-dependence built into the mod. UNIFAC (Do.) method.

In this case, the experimental temperature is used to generate one of the predicted datasets, and two additional datasets are generated 33.3 K below and 33.3 K above this value, while the fourth dataset is generated 66.6 K above the experimental temperature. The entire composition range is then incremented using 10 equal molar steps at each temperature, resulting in a total of 44 data points covering a temperature range of 100 K for each DDB-VLE dataset evaluated.

These four pseudo datasets were then simultaneously regressed using the NRTL equation. The component-based binary interaction parameters of the NRTL equation were fitted using a quadratic relationship in temperature (Equation (5-13)), so that the temperature-dependence *built into* the mod. UNIFAC (Do.) method could be captured:

$$\Delta u_{ij} = a_{ij} + b_{ij}(T) + c_{ij}(T)^2 \tag{5-13}$$

So compared to the 3 parameters required to fit the standard form of the NRTL equation to an isothermal dataset (recall Chapter 3.2.2), a total of 7 parameters are being fitted to the predicted results (including the nonrandomness parameter α). Using the Simplex Nelder-Mead method (Nelder, et al., 1965) to minimize the average of the relative deviations in pressure squared (Pdev^2, calculated using Equation (5-1)) between the predictions and the calculated values. The regressed predictions and fitted parameters were then used to calculate the experimental DDB-VLE datasets.

5.4.1. **Calculations**

Some simulators by default use the predicted activity coefficients at infinite dilution to fit missing binary-interaction parameters (e.g. AspenPlus); however given the errors observed in predicting these values (recall Chapters 5.2 and 5.3), this approach is likely prone to additional errors.

A recent journal article (Cadoret, et al., 2009) will be used to illustrate this point, where binary parameters were estimated based on activity coefficients at infinite dilution and constant temperature predicted using mod. UNIFAC (Do.) with the help of AspenPlus (using a temperature of 298.15 Kelvin by default). A case was made against the GC2gE conversion (group contribution to g^E-model conversion) of estimated activity coefficients at infinite dilution of mod. UNIFAC (Do.) in favour of regressing the predictions of COSMO-SAC (Lin, et al., 2002) over the entire composition range.

In both situations the NRTL equation was used; however it was observed that parameters directly calculated from predicted activity coefficients at infinite dilution for the system ammonia (1) + tert-butanol (2) using mod. UNIFAC (Do.) (NRTL GC2gE (PCES) of Figure 32) do not reproduce the group contribution model results well compared to those regressed using COSMSO-SAC over the entire composition range (NRTL GC2gE (COSMO) of Figure 32).

The authors acknowledge that it has been observed in literature (Athés, et al., 2008) that results can be sensitive to activity coefficients at infinite dilution, and simply state that the estimation of missing binary parameters using predicted activity coefficients at infinite dilution must be “handled with care.” Cadoret, et al. use this as justification for making the conclusion that COSMO-SAC is more suitable than mod. UNIFAC (Do.) for the specific mixture in question (ammonia/tert-butanol), and that the COSMO-SAC model can be used as a suitable alternative to the limitations of UNIFAC-type methods (e.g. lack of functional groups, inaccurate estimations).

Although the later statement rings true (i.e. use the best tool for the job), the former statement is based on an "apples-to-oranges" comparison; meaning that in one case a known model limitation and weakness was used to fit the binary parameters of the NRTL equation, while the entire composition range of COSMO-SAC predictions were used in the other (which essentially deweights the errors in the infinite dilute region in comparison).

If binary parameters for the NRTL equation are fitted to the predictions of mod. UNIFAC (Do.) over the entire composition range (NRTL GC2Ge (USER) of Figure 32), however, the regressed results qualitatively match those obtained by COSMO-SAC (as is depicted in Figure 32). Therefore predicitons over the entire composition range should be used to estimate missing binary parameters for component-based methods such as NRTL; since otherwise, the results obtained by only using activity coefficients at infinite dilution might lead to erroneous results (as shown in Figure 32).

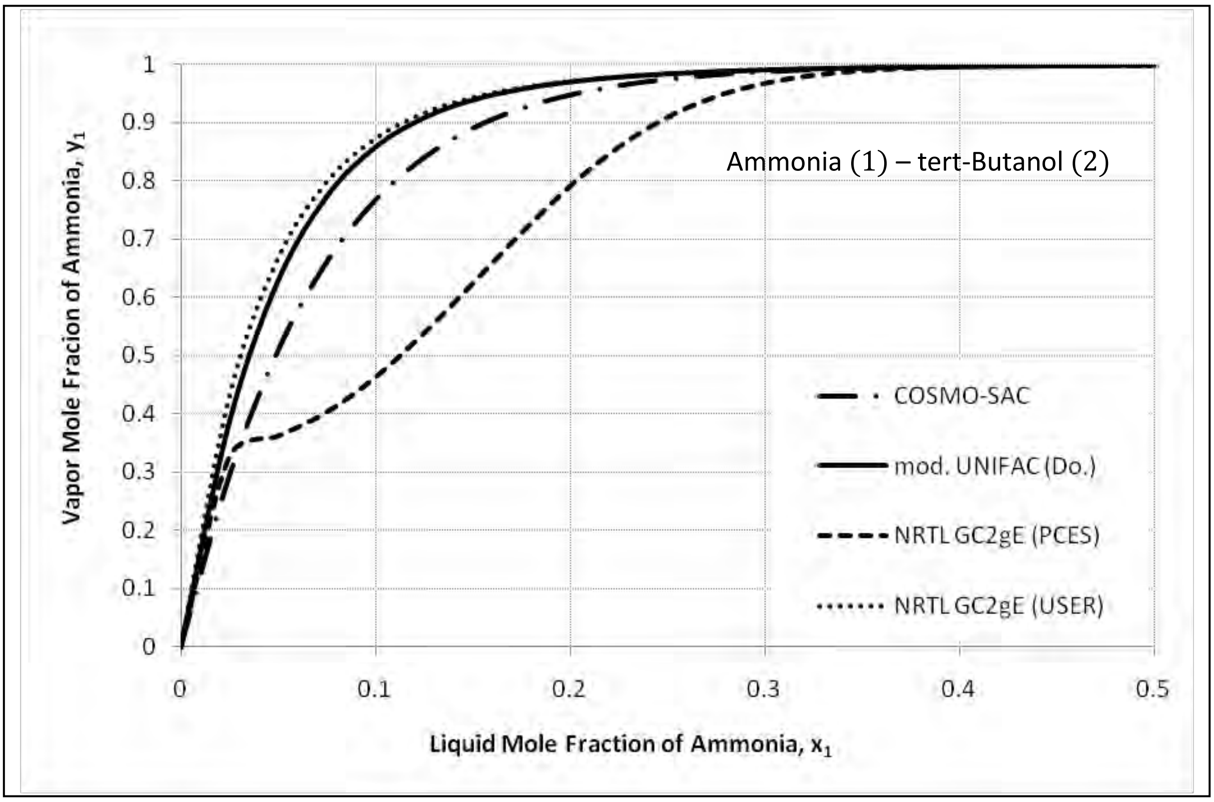
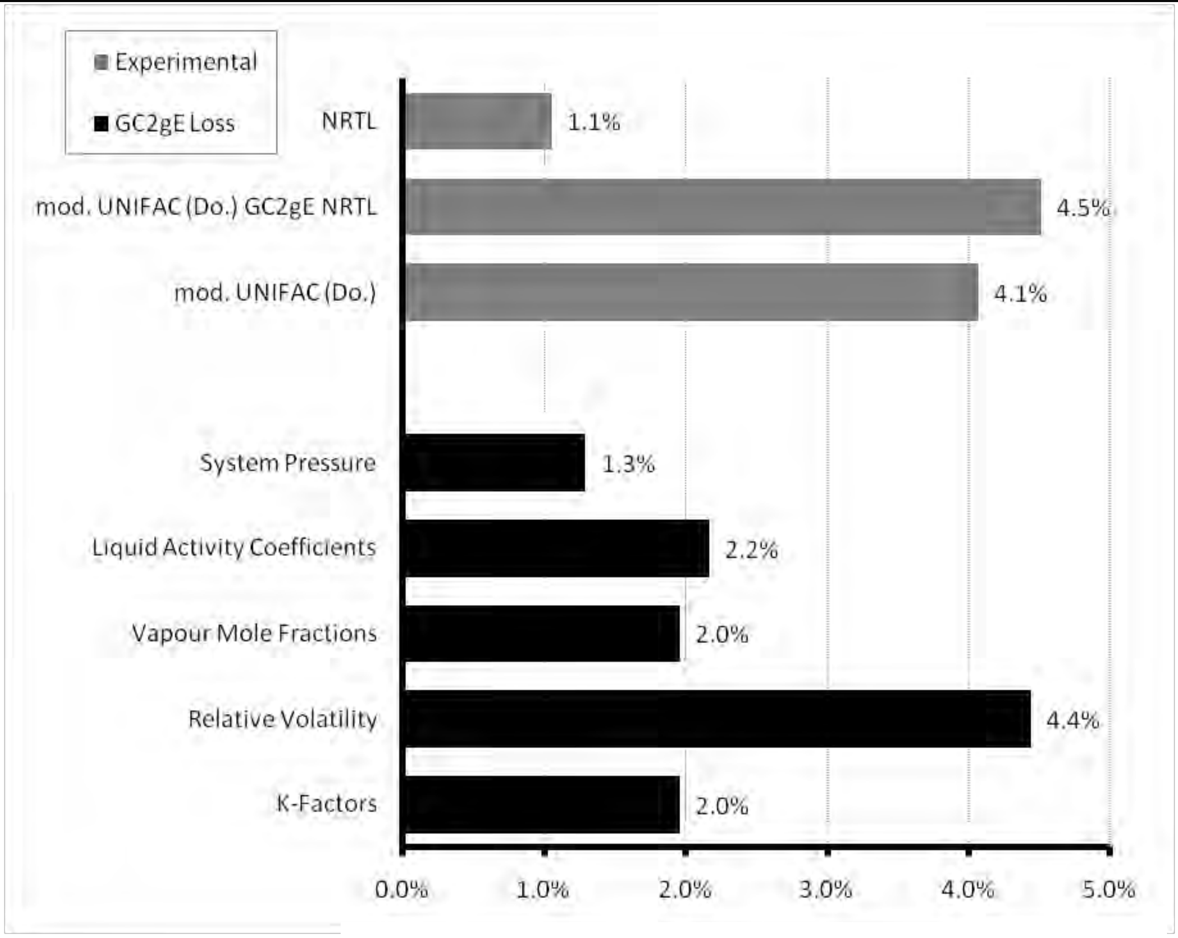


Figure 32 AspenPlus was used to recreate Figrue 2(b) in the publication of Cadoret, et al. 2009, where parameter estimations using predictions were made for the system ammonia (1) + tert-butanol (2) at a pressure of 0.5 atm. The lines labled as COSMO-SAC and mod. UNIFAC (Do.) represent the predictions using these methods, and NRTL GC2gE (PCES) the AspenPlus default estimation of missing binary parameters (using mod. UNIFAC (Do.)) and NRTL GC2Ge (USER) representing the regression of predictions over the entire composition range.

The NRTL equation was therefore fitted to the predictive results of mod. UNIFAC (Do.) over the entire composition range at four different temperatures, where the binary-parameters were fitted using a quadratic dependence in temperature (see Equation (5-13)); furthermore, the nonrandomness factor (α) was also used as a fitting constant, where the value was kept between 0.1 and 0.4 (within a range of physically significant values typically encountered in practice). The fitting of the 7 model parameters per dataset gives the NRTL equation maximum flexibility, and should represent the best-case scenario for translating the predictions/knowledge of mod. UNFAC (Do.) into the component-based model of NRTL (i.e. in performing the GC2gE conversion).

After the NRTL parameters were fitted to the predictions of mod. UNIFAC (Do.), they were then used to recalculate the data points of the selected DDB-VLE systems satisfying the criteria already discussed in the beginning of this Chapter, and then the results were compared to the predicted values for these same points using mod. UNIFAC (Do.). The resulting relative average deviations (RAD) between predicted values obtained from the mod. UNIFAC (Do.) method and the calculated values obtained from the NRTL equation fitted to these predictions are presented in Figure 33 below (shown as the black-set of bars).

The only values which can be related to actual experimental data, however, are the pressure values (recall only Px(T) datasets were used for evaluation purposes). Therefore, Figure 33 also includes the resulting RAD values between the calculated model-results and the experimental pressure (shown as the grey-set of bars) for mod. UNIFAC (Do.) and the NRTL equation used in the GC2gE conversion with parameters fitted using a quadratic dependence in temperature. Also included for comparison are the results of the NRTL equation (obtained in Chapter 5.1) using parameters fitted directly to each experimental dataset.



Relative Average Deviations:

$$\Delta Property_{RAD} = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\left(Property_{i,calc} - Property_{i,exp}\right)}{Property_{i,exp}^{\infty}}\right]$$

Figure 33 Relative average deviations (RAD) between predicted values obtained from mod. UNIFAC (Do.) and calculated values obtained from the NRTL equation, using parameters (quadratic in temperature) fitted to these same predictions, for select isothermal DDB-VLE systems.

As can be seen in Figure 33 (note the black-bars), a noticeable amount of deviation exists between the predicted values and the results of the NRTL equation; however one must also keep in mind that the calculation of relative-average-deviations is strongly dependent on the number of data points used. Although the RAD in pressure is only 1.3%, for instance, the value is only semi-quantitative (i.e. in some cases it could be much higher, and in other cases it could be close to zero). Some may say this kind of comparison is a moot-point since these values (the black-bars of Figure 33) are calculated relative to the predictions of mod. UNIFAC (Do.); however, in practical applications where the GC2gE conversion is used, experimental information is not known.

Differences may arise from the inability of the NRTL equation (or another g^E-model used) to be fitted to certain shapes predicted by the group contribution method used in the process of performing a GC2gE conversion (whether the predicted description is correct or not). Although no examples falling into this category were found in the present study, it is still nonetheless a possibility. The RAD in pressure of 1.3%, therefore, is likely due to the fitting procedure of regressing pseudo data at multiple temperatures in this case (needed for the fitting of NRTL parameters quadratic in temperature).

An example illustrating this observation is shown below in Figure 34, where the predictions of mod. UNIFAC (Do.) are compared against the calculations using the fitted results from the GC2gE conversion using the NRTL equation. For this particular system, the RAD in pressure (relative to the predictions) is 34.6%. Given the ideal nature of the predictions, it is hard to believe that this type of behaviour cannot be fitted using the NRTL equation. The reason for the unfavoured fit is undoubtedly caused by attempting to capture the temperature-dependence of the mod. UNIFAC (Do.) method. This assertion was verified by performing another GC2gE conversion using the NRTL equation, but this time fitted at the experimental temperature of the system only (the series is marked by a dashed line in Figure 34).

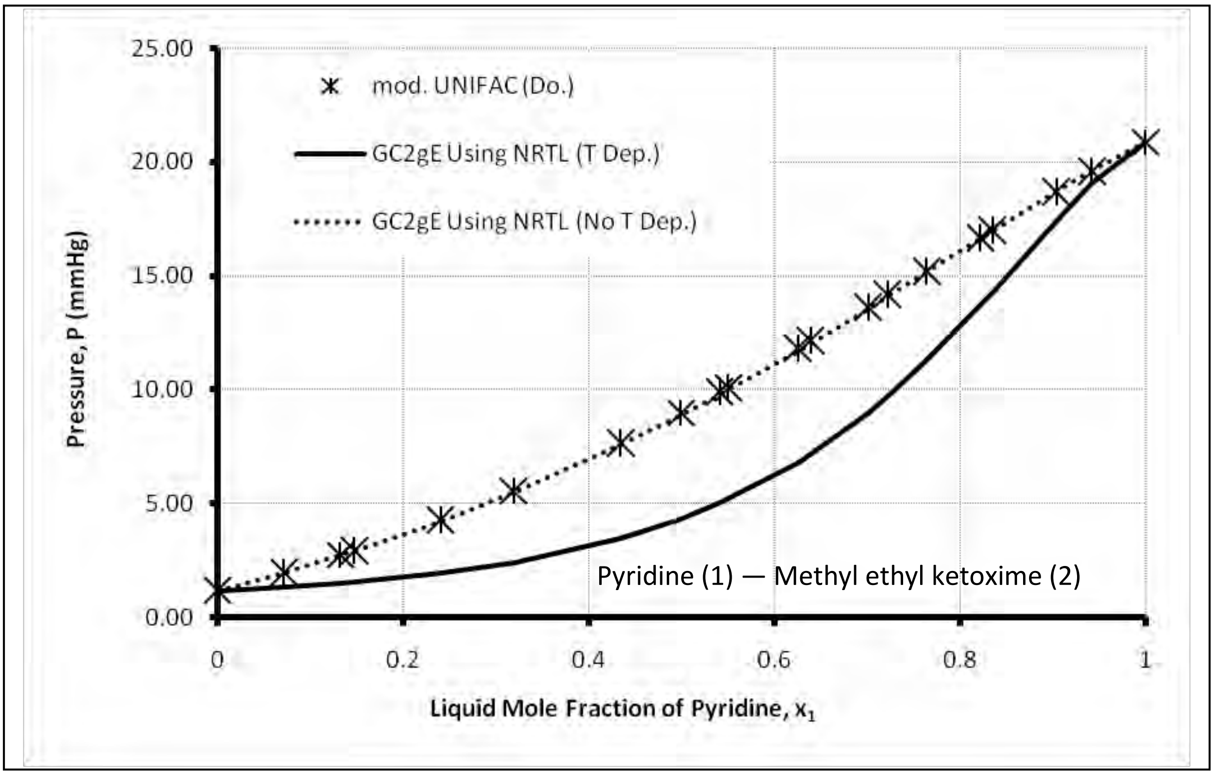


Figure 34 DDB-VLE dataset number 7697, Px(T) data, consisting a mixture of pyridine (1) + methyl ethyl ketoxime (2) at 298.15 K. Depicts the "loss in translation" which occurs from a GC2gE conversion using the NRTL equation, where parameters quadratic in temperature were fitted.

Although high RAD values are obtained for the GC2gE conversion process used in this Chapter (the black-bars of Figure 33), these discrepancies vanish when the calculated results using the fitted results are compared against the *actual* experimental data (the grey-bars of Figure 33). The differences between the calculated results using the mod. UNIFAC (Do.) method, and the results obtained using the NRTL equation fitted to these same predictions, is only 0.4% (the 4.5% of the GC2gE conversion using the NRTL equation minus the -4.1% of the mod. UNIFAC (Do.) method). This small deviation of 0.4% is well within the experimental uncertainty of mod. UNIFAC (Do.), and may attribute to why the "loss in translation" effect has gone largely unnoticed, or at least not officially addressed in open literature. This is especially evident when a distribution of the difference in Pdev^2

values between the GC2gE conversion results and the results of the mod. UNIFAC (Do.) method (see Figure 35 below), where only about a 1% of the datasets evaluated have an absolute difference greater than 0.0001 in Pdev^2.

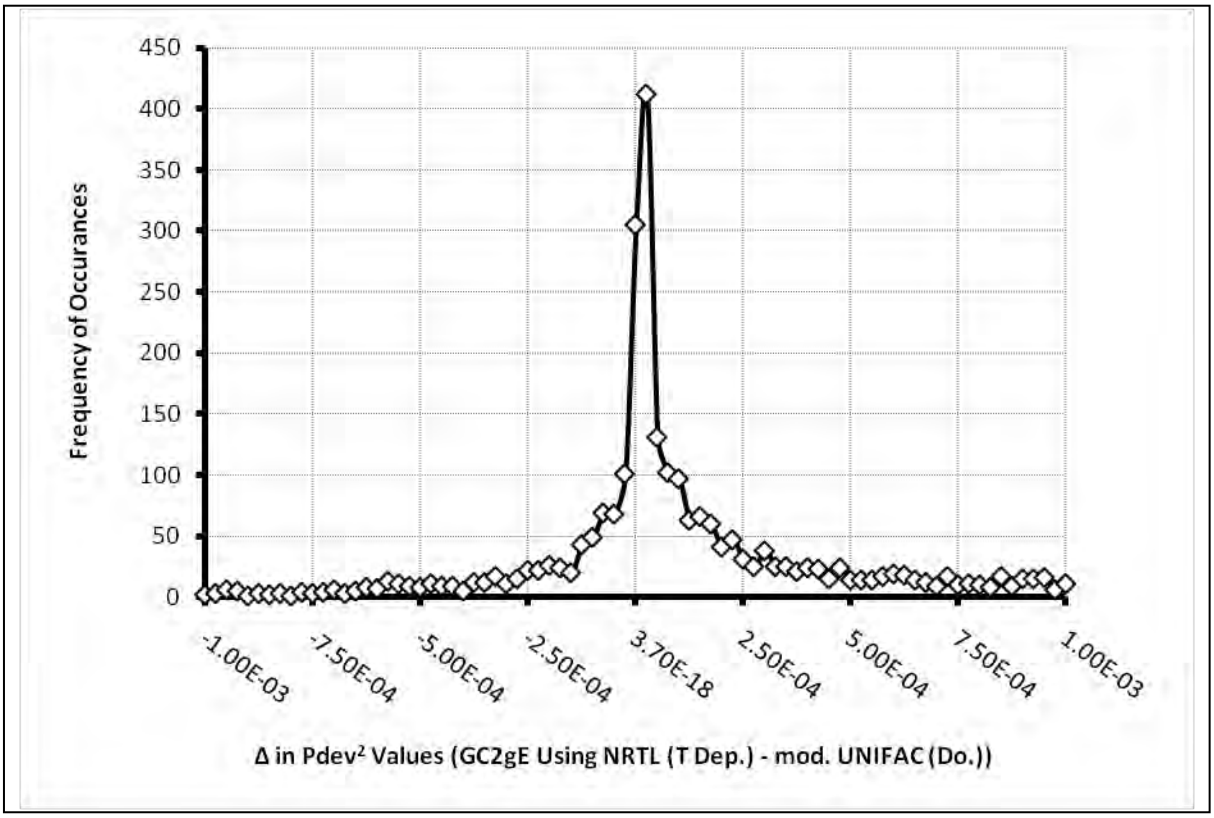


Figure 35 Distribution of the difference in Pdev^2 values between the GC2gE conversion results using the NRTL equation (fitting parameters quadratic in temperature) and the results of the mod. UNIFAC (Do.) method.

Since the extent of this "loss in translation" will be unique for each individual system, and given that practical applications are rarely limited to component pairs, the system of acetone (1) + benzene (2) + methanol (3) at 298.15 K will be used as a case-specific example to illustrate the potential effects for a multicomponent mixture (see Figure 36).

Figure 36 shows a comparison of DDB-ACM experimental data against the predictions using mod. UNIFAC (Do.) and the results obtained from the predictions of mod. UNIFAC (Do.) regressed by the NRTL equation. Although mod. UNIFAC (Do.) represents the experimental activity coefficients of methanol at infinite dilution in a mixture of acetone and benzene quite well, the GC2gE conversion using the NRTL equation results in a poorer representation of the experimental data—a "loss in translation" has occurred.

For completeness, a comparison of predictions made with the mod. UNIFAC (Do.) method and the calculated results obtained from the GC2gE conversion using the NRTL equation is shown for the systems acetone (1) + methanol (2) (see Figure 37) and benzene (1) +methanol (2) separately (see Figure 38). In both cases, the GC2gE conversion using the NRTL equation fails to reproduce the raw predictions made using the mod. UNIFAC (Do.) equation, resulting in a worse representation of the real fluid behaviour for both systems. The largest of these negative deviations are seen for the binary system benzene (1) + methanol (2) (Figure 38), and appear to be the main binary-pair which is influencing the large negative deviations calculated for the acetone (1) + benzene (2) + methanol (3) system of Figure 36; in this case, the mod. UNIFAC (Do.) method gives the best representation of the real phase behaviour for the ternary system considered.

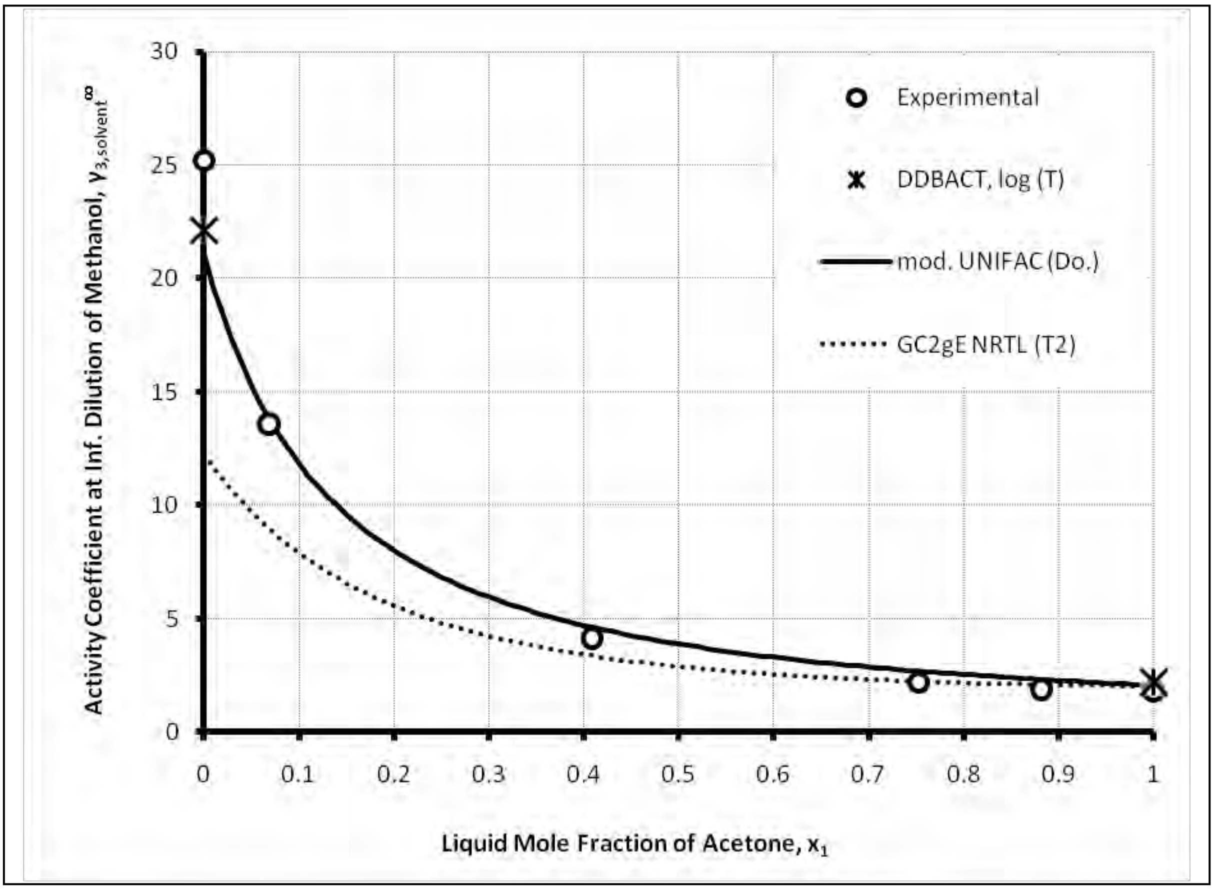


Figure 36 Comparison of mod. UNIFAC (Do.) predictions and calculations made with regressed predictions using the NRTL equation (fitted using a quadratic dependence in temperature) against experimental dataset DDB-ACM 647 (methanol (3) in acetone (1) + benzene (2) mixture at 298.15 K). Included in the comparison are estimates made using an inverse temperature log-fit to relevant experimental data from DDB-ACT—methanol in acetone (see Figure 37(b)) and methanol in benzene (see Figure 38(b)).

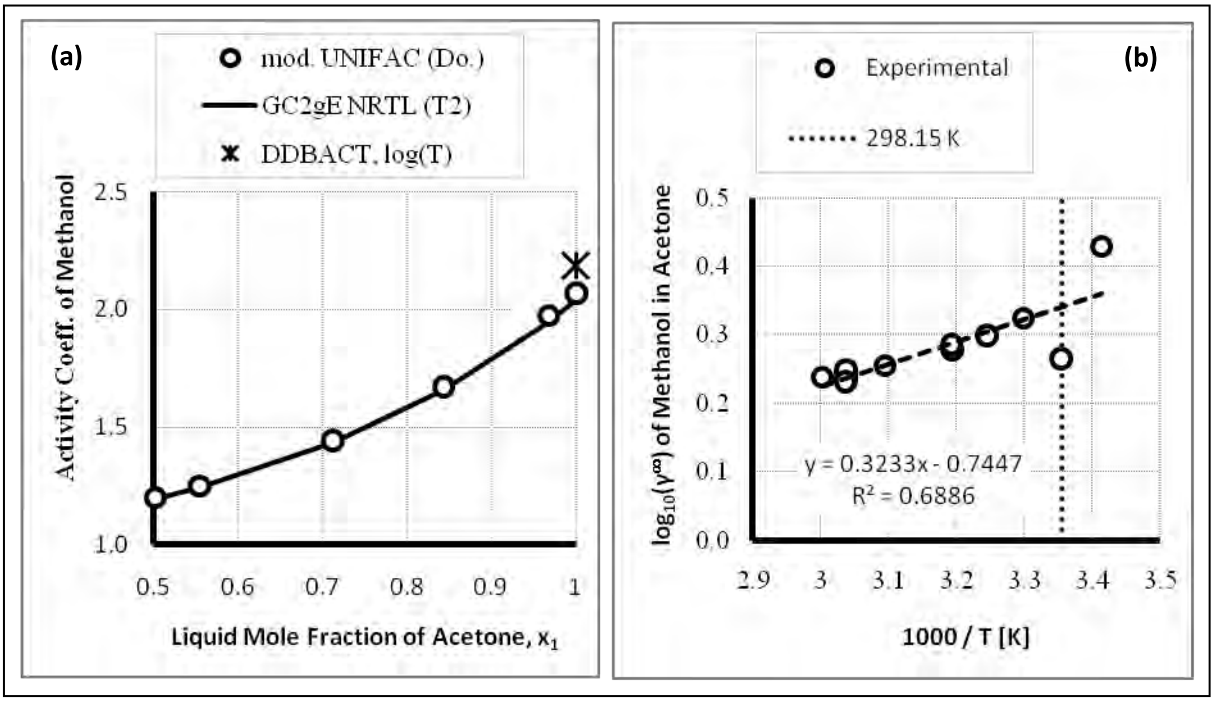


Figure 37 Part (a) (corresponding to DDB-VLE set 19175) shows mod. UNIFAC (Do.) predictions for the system acetone (1) + methanol (2) at 298.15 K against the calculated results of the NRTL equation fitted to predicted datasets generated using mod. UNIFAC (Do.). Included in the comparison is the estimated activity coefficient at infinite dilution, obtained from a linear-log-fit of experimental data from DDB-ACT (see part (b)).

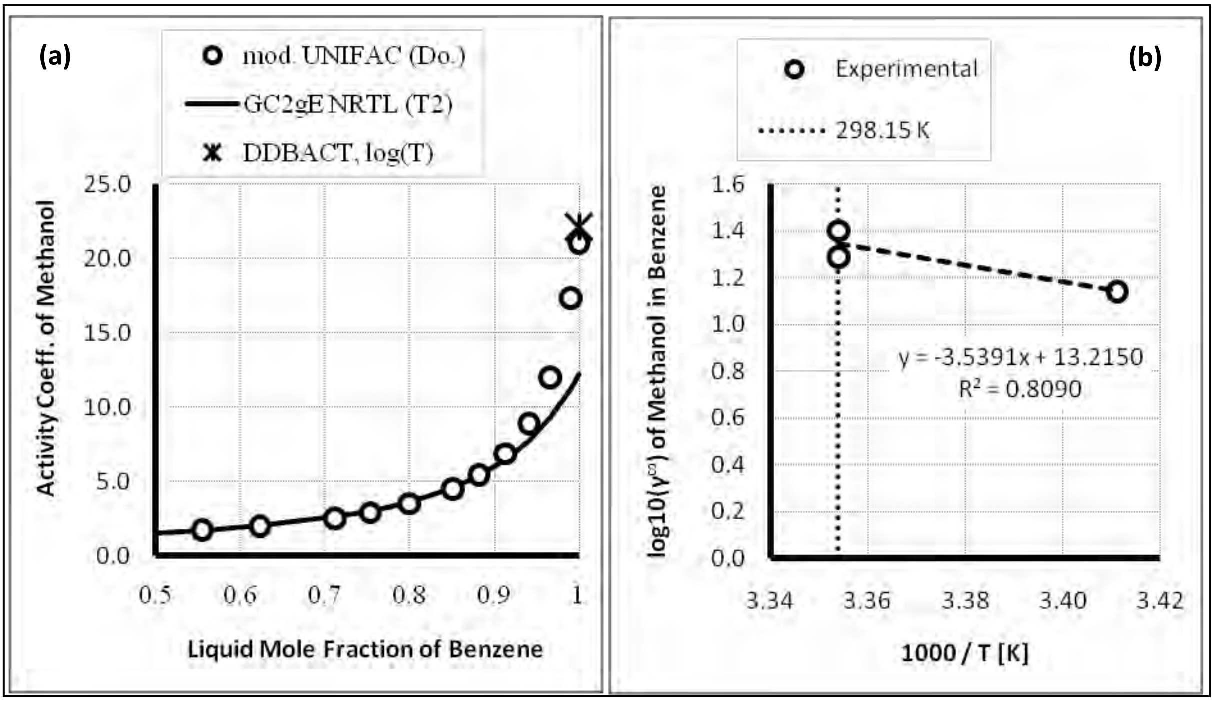


Figure 38 Part (a) (corresponding to DDB-VLE set 5008) shows mod. UNIFAC (Do.) predictions for the system benzene (1) + methanol (2) at 298.15 K against the calculated results of the NRTL equation fitted to predicted datasets generated using mod. UNIFAC (Do.). Included in the comparison is the estimated activity coefficient at infinite dilution, obtained from a linear-log-fit of experimental data from DDB-ACT (see part (b)).

It should be noted that regressing NRTL binary interaction parameters to predicted pressures over the whole concentration range is by no means the only possibility to derive the NRTL parameters from predictive methods such as mod. UNIFAC (Do.). One may use, for example, predicted activity coefficients at infinite dilution or activity coefficients over the whole concentration range (recall Figure 32), or one may predict a complete VLE dataset (x, y, P, (T) — data) and regress any other property of interest. For example predicted separation factors can be fitted; strongly weighting values close to unity, as they represent the largest separation effort in unit operations such as distillation.

This seems to be a good choice at first sight, but it is easy to find examples where this procedure is less advisable. For example, the systems of benzene (1) + NMP (3) and cyclohexane (2) + NMP (3) exhibit separation factors far from unity, especially at higher concentrations of the high boiling component NMP. The most important application of parameters for these systems is the extractive distillation of benzene (1) and cyclohexane (2) using NMP (3) as an entrainer.

NMP is fed into the column a little below the top stage and flows with the liquid down into the bottom. Its concentration in the liquid is typically well above 50 mol%. The important property for the column is the separation factor between benzene (1) and cyclohexane (2); α_{12} at high NMP concentrations that can be derived from the ratio of α_{13} to α_{23}, which are both far from unity. Errors in α_{13} or α_{23} therefore strongly affect the column design (recall Figure 23 of the preceding Chapter).

Many further examples of this kind can be found and it is generally a good idea to regress the activity coefficients generated by the predictive mixture model, so that important calculated properties required for process simulation can be closely matched using the fitted model. Since practitioners often make use of process simulators such as AspenPlus to perform GC2gE conversions like this, a further example is shown in Figure 39 below. Therefore, group contribution methods should not be used as second-choice data generators because this can result in a “loss in translation,” and may cause additional errors to be introduced into the simulations of practitioners.

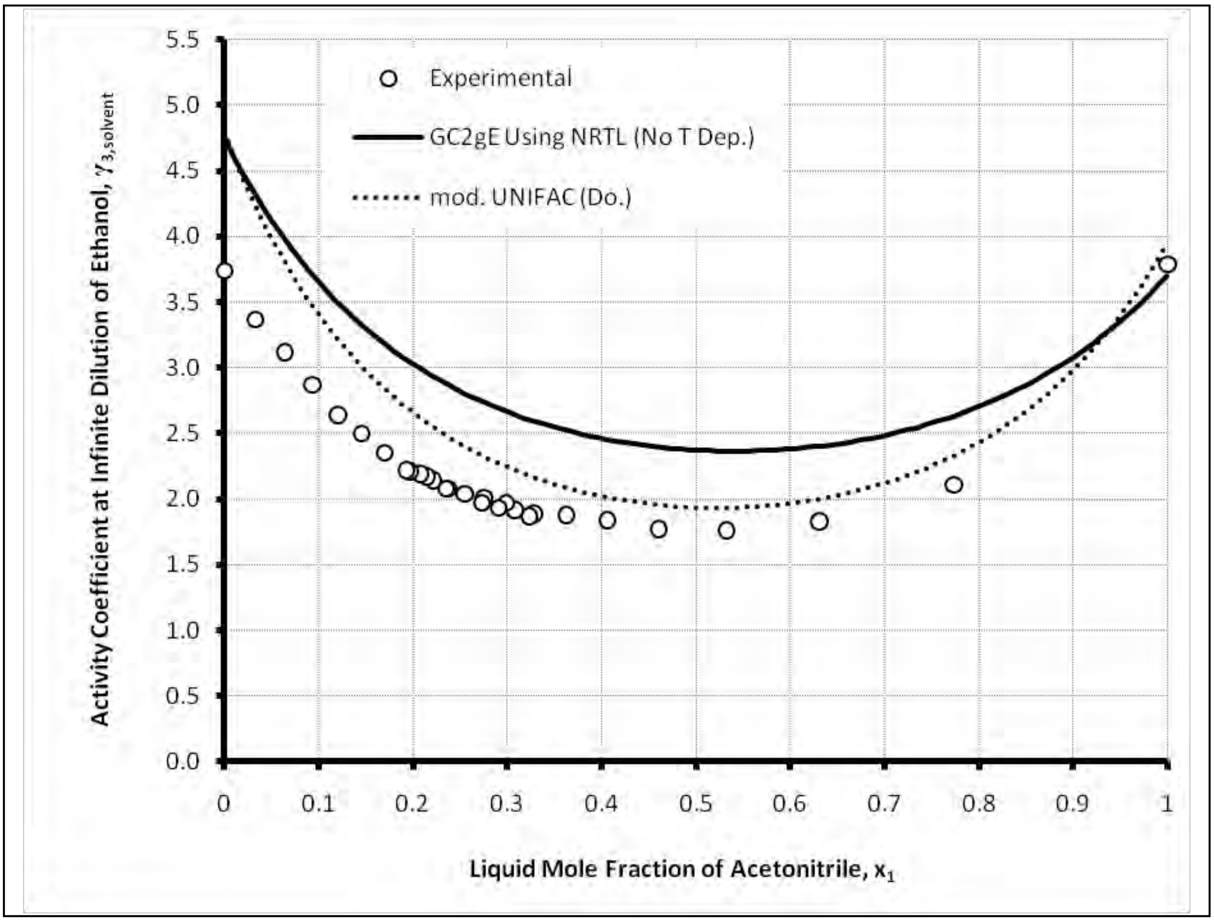


Figure 39 DDB-ACM dataset number 731, consisting of the mixture acetonitrile (1) + water (2) + ethanol (3) at 298.15 K. Depicts the "loss in translation" which occurs when a GC2gE conversion using the NRTL equation is performed within the AspenPlus simulator environment, where the NRTL parameters (not quadratic in temperature) were fitted to predicted activity coefficients over the entire concentration range obtained using the mod. UNIFAC (Do.) method.

5.5. Proposed Hybrid Method

The coexistence of different models for the description of the real liquid phase behaviour is evidence that not any one method is overwhelmingly superior; as a result, practitioners may use more than one model to obtain an accurate representation of their system (e.g. the process simulator AspenPlus allows to choose a model for each unit operation-block separately). All of these models require the fitting of model parameters to experimental data or results of predictive methods. For component-based models like the NRTL equation, this requires the fitting of binary-interaction parameters for every possible pair of components (in order to be considered fully defined). Using group contribution methods as “additional data sources” enables practitioners to utilize any measured data they may have available, where the adjustment of specific component binaries to “actual” experimental data (i.e. not generated using a group contribution method) typically leads to a better results.

As discussed in the previous chapter, however, the regression of predicted results can potentially lead to a “loss in translation” when moving from the group contribution approach into the component-based approach (i.e. when performing the GC2gE conversion). One way to completely remove this effect would be to use group predictions directly in process simulations, which is typically restricted to cases where a very large number of components are present in the mixture (cases in which component-based models would require an excessively large number of parameters). This would mean, however, that the calculated behaviour of specific component binaries could no longer be influenced independently of each other for reasons discussed in Chapter 5.1 (Evaluation of the Solution of Groups Concept). Additionally the observed deviations between predicted results and available experimental data discussed in Chapter 5.2 (Ability to Predict Activity Coefficients) and Chapter 5.3 (Ability to Represent Multicomponent Mixtures) could no longer be easily corrected for.

In order to meet industrial needs a new method is proposed here, which should enhance the usability of group methods such as mod. UNIFAC (Do.) without introducing unnecessary uncertainty into process simulation. To this end a “hybrid group contribution model” has been developed that, besides group interactions, employs component interactions adjustable to actual experimental data. As briefly stated in Chapter 4 (Proposed Method), the proposed hybrid model has the following general form (restated here for ease of reading):

$$\ln\gamma_i = \overbrace{\underbrace{\ln\gamma_i^C}_{\substack{Combinatorial\\contribution}} + \underbrace{\ln\gamma_i^R}_{\substack{Residual\\contribution}}}^{base-GC\ model} + (\text{LC-Model}) = \ln\gamma_i^{GC} + \ln\gamma_i^{LC} \qquad (4\text{-}1)$$

where a component-based correction is layered on top of a group contribution model such as mod. UNIFAC (Do.), essentially creating a layer-cake method. This simple augmentation/methodology could conceivably be extended to other situations, where the combination of different models may result in a better description of the real behaviour of mixtures (or an increase of range of application) than simply using the individual base methods making up the layers of the hybrid model. In a number of cases the hybrid approach might even provide a better representation of experimental data than the component-based model alone; since the model adjusts to the predictions of the group contribution model employed (and not from Raoult’s Law). Therefore if the group contribution layer could describe 95% of the real mixture behaviour, for example, then the component-based layer would only need to provide the last 5% (i.e. each layer of the hybrid model shares the load; recall Figure 6).

In fact, similar ideas are employed every day in performing practical phase equilibrium calculations; the “Gamma-Phi” methodology is one such case. Where an equation of state (EOS) is used for defining the fugacity coefficient for the description of the vapour phase, and where an activity coefficient model is used to define activity coefficients used for the description of the liquid phase. For low to moderate pressures, this arrangement typically results in the best description of the real phase behaviour of mixtures. The non-idealities in the liquid phase are handled by an activity coefficient model (where many mixing rules of/and equations of state typically have trouble) and the non-idealities in the vapour phase are handled by an equation of state (where activity coefficients do not account for).

In order to understand how well the hybrid model/approach represents the real behaviour of mixtures, hybrid models were fitted to experimental data and then compared against the individual fitting results of the component-based models employed. Two hybrid variations were tested; one that uses the NRTL equation and another that uses the Wilson equation for the component-based correction[10]. Both hybrid variations use the mod. UNIFAC (Do.) method as the base GC-layer, since it tends to give superior results in most cases compared to the UNIFAC method (recall Chapter 5.2 and Chapter 5.3; also in literature (Gmehling, 2009)). For ease of reference these hybrid models will be referred to as NRTL-FAC(Do) and Wilson-FAC(Do) respectively.

Binary isothermal data—Px(T)—satisfying the criteria of Chapter 5.1.2 were used for evaluation purposes, where NRTL-FAC(Do), Wilson-FAC(Do), NRTL, and Wilson equations were fitted to each dataset individually, using the Simplex Nelder-Mead method to minimize the mean relative deviation squared in pressure (the objective function, O.F.) for each dataset (recall Equation (5-1)). A illustration of this procedure, and of the method in general, is provided in Figure 40 below.

---

[10] The formalism can also be applied to combine, as an example, the UNIQUAC equation with mod. UNIFAC (Do.), but it would likely be a good idea to use only the residual part of UNIQUAC as the combinatorial contribution is already included in the group contribution model.

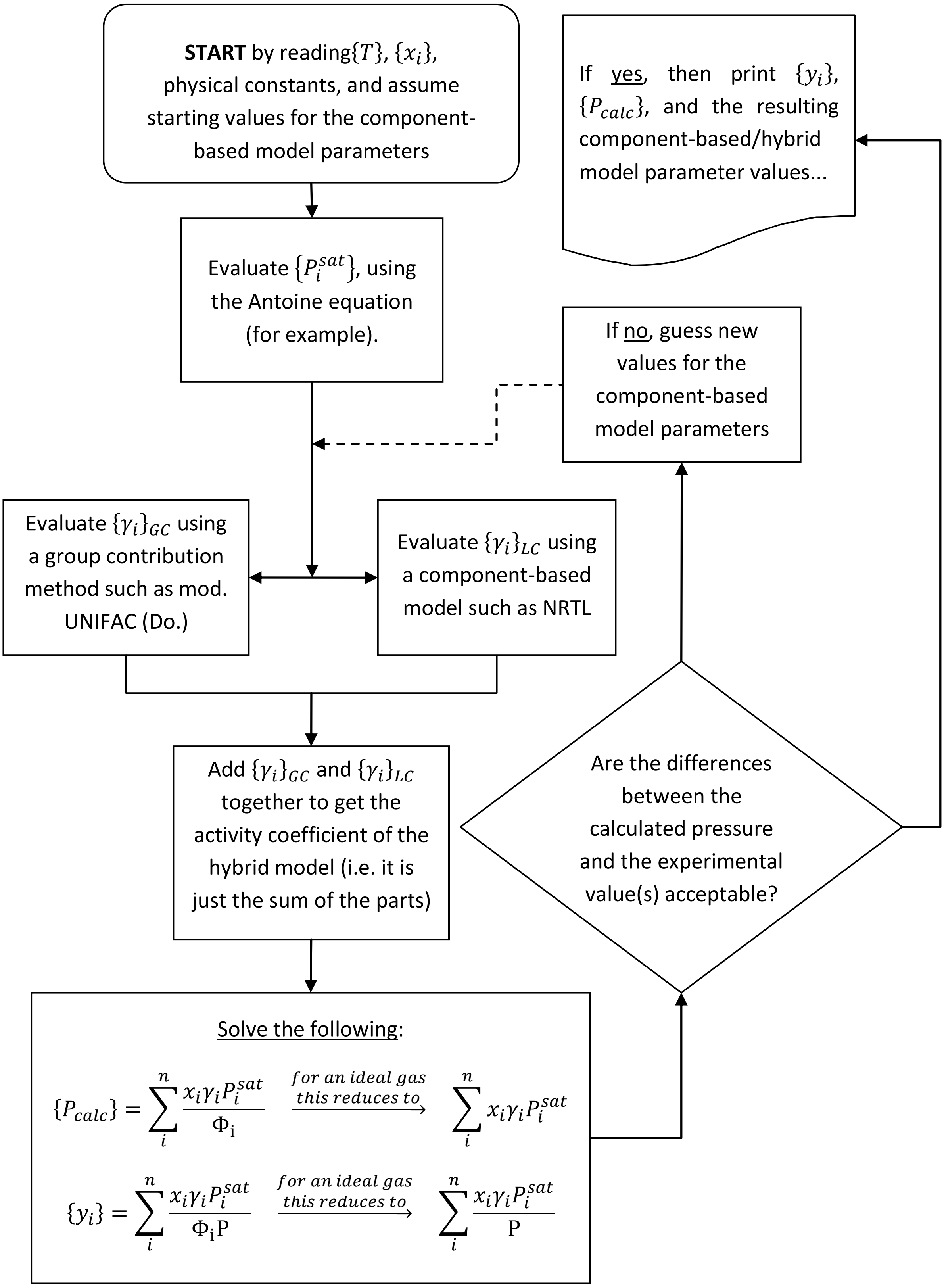


Figure 40 Illustrates the implementation of the proposed hybrid methodology, and the fundamental calculation steps used to determine the hybrid model parameters.

5.5.1. Calculations

In a likewise fashion to Chapter 5.1.2, ratios of Pdev^2 values for comparable models (calculated using Equation (5-1)) were used to visualize the goodness-of-fit for the NRTL-FAC(Do.) and Wilson-FAC(Do.) hybrid models. The calculated Pdev^2 values for each hybrid model were compared against the results of the component-based model used in the tuning layer, the predicted results of mod. UNIFAC (Do.), and the opposing hybrid model. In all cases a minimum relative squared deviation of 10^{-6} was used when calculating the ratios (since deviations of $\sqrt{10^{-6}} = 0.001$ are mostly acceptable in practical applications). These comparisons are presented in Figure 41 below, where comparisons are shown in the part (a) of Figure 41 for the regression results using the NRTL-FAC(Do.) model, and those of the Wilson-FAC(Do.) model are shown in part (b) of Figure 41.

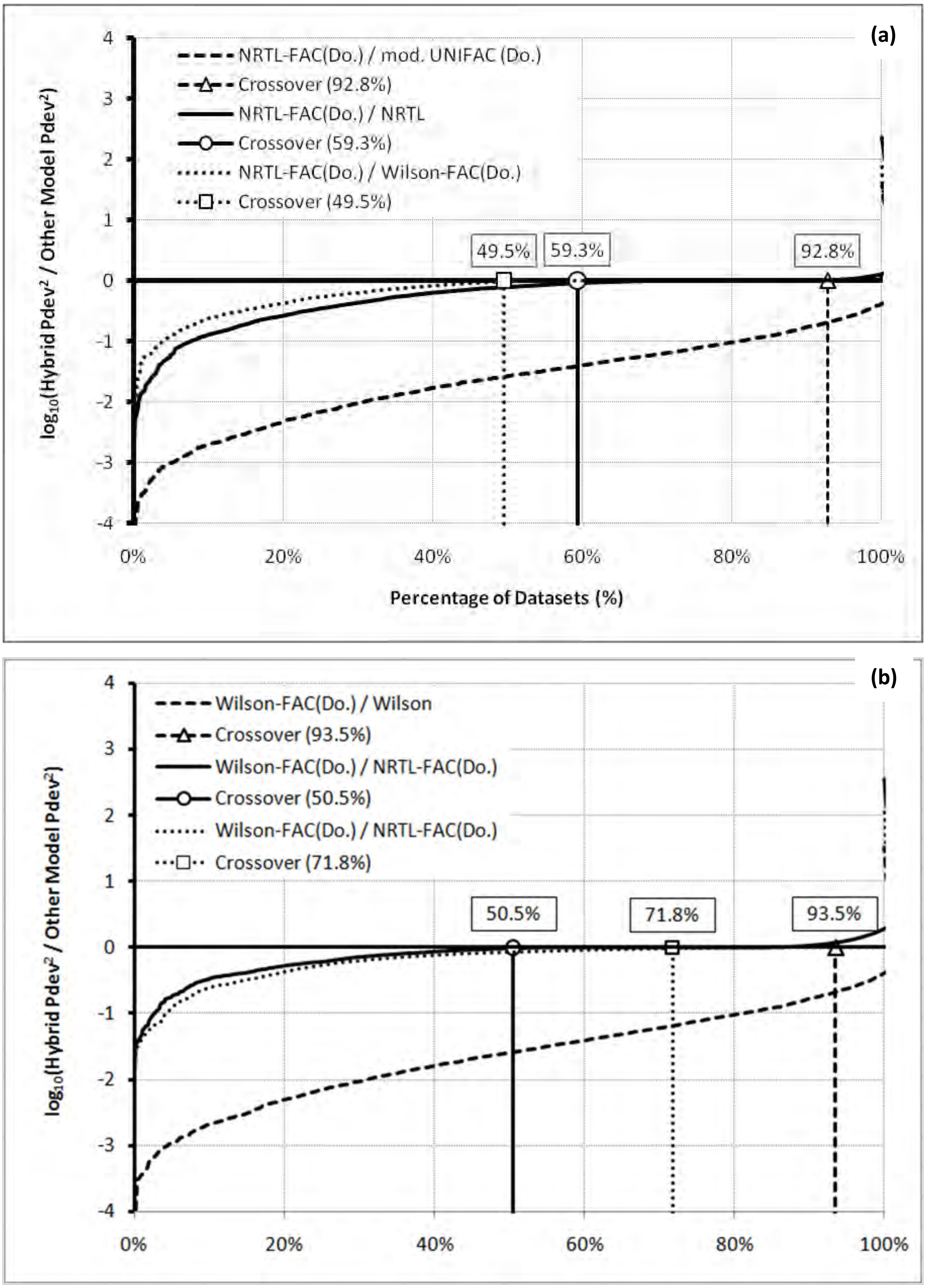


Figure 41 Log$_{10}$ ratio of Pdev^2 values (Equation (5-1)), where the fitting results of the hybrid models NRTL-FAC(Do) and Wilson-FAC(Do) are compared against the fitting results obtained using the NRTL and Wilson equations respectively. For completeness, the fitting results of the NRTL-FAC(Do) hybrid model (numerator) are compared against the fitting results obtained using the Wilson-FAC(Do) hybrid (denominator).

Since the ordinates of Figure 41 are $\log_{10}$, any values less than y = 0 then represent cases in which the hybrid model reproduces the experimental values of the dataset better than the regressions using the component-based model used for the tuning-layer of the method. The term "crossover," in this case, is then used to represent the point at which each line crosses *over* the y = 0 line. This then quantifies the percentage of datasets where the regressions using the hybrid model outperform, or are equivalent to, the regression results of the model being compared (the numerator of the ratio for each curve). These values have been summarized in Table 9 below for convenience.

Dataseries Name			Crossover Point	
NRTL-FAC(Do.)	/	mod. UNIFAC (Do.)	92.8 %	(3571/3848)
NRTL-FAC(Do.)	/	NRTL	59.3 %	(2283/3848)
NRTL-FAC(Do.)	/	Wilson-FAC(Do.)	49.5 %	(1906/3848)
Wilson-FAC(Do.)	/	mod. UNIFAC (Do.)	93.5 %	(3596/3848)
Wilson-FAC(Do.)	/	Wilson	50.5 %	(1945/3848)
Wilson-FAC(Do.)	/	NRTL-FAC(Do.)	71.8 %	(2761/3848)

Table 9 The percentage of data at the crossover point for each line in Figure 41, where the number of datasets leading to these percentages are given in the parentheses to the right of the crossover value.

As can be seen from Figure 41 and Table 9 above, both hybrid models successfully influence the calculated behaviour of mod. UNIFAC (Do.). This is very evident by the large crossover values for both curves, where NRTL-FAC(Do.) compared to the mod. UNIFAC (Do.) method results in a crossover value of 92.8%, and the Wilson-FAC(Do.) model having a crossover value of 93.5% in comparison. This means that the predictions of mod. UNIFAC (Do.) are better than the NRTL-FAC(Do.) model 7.3% of the time, and better than Wilson-FAC(Do.) 6.4% of the time; however when these cases are analyzed closer, they are all datasets where the mod. UNIFAC (Do.) method provides good agreement with the experimental data points. The average of the absolute differences between the Pdev^2 values of NRTL-FAC(Do.) and the Pdev^2 values using the mod. UNIFAC (Do.) method for these cases is only 9.72E-04 (with a maximum absolution deviation of 1.25E-02), and only 9.64E-04 for the Wilson-FAC(Do.) equation (with a maximum absolute deviation of 4.64E-02). These differences can then be attributed to the fitting program, which likely had trouble reconciling the small deviations with the step values used in the Simplex Nelder-Mead for fitting the model parameters. An illustration is provided for the system acetone (1) + water (2) below in Figure 42, where the absolute difference between the Pdev^2 values using the mod. UNIFAC (Do.) method and the hybrid models is roughly 1.03E-02.

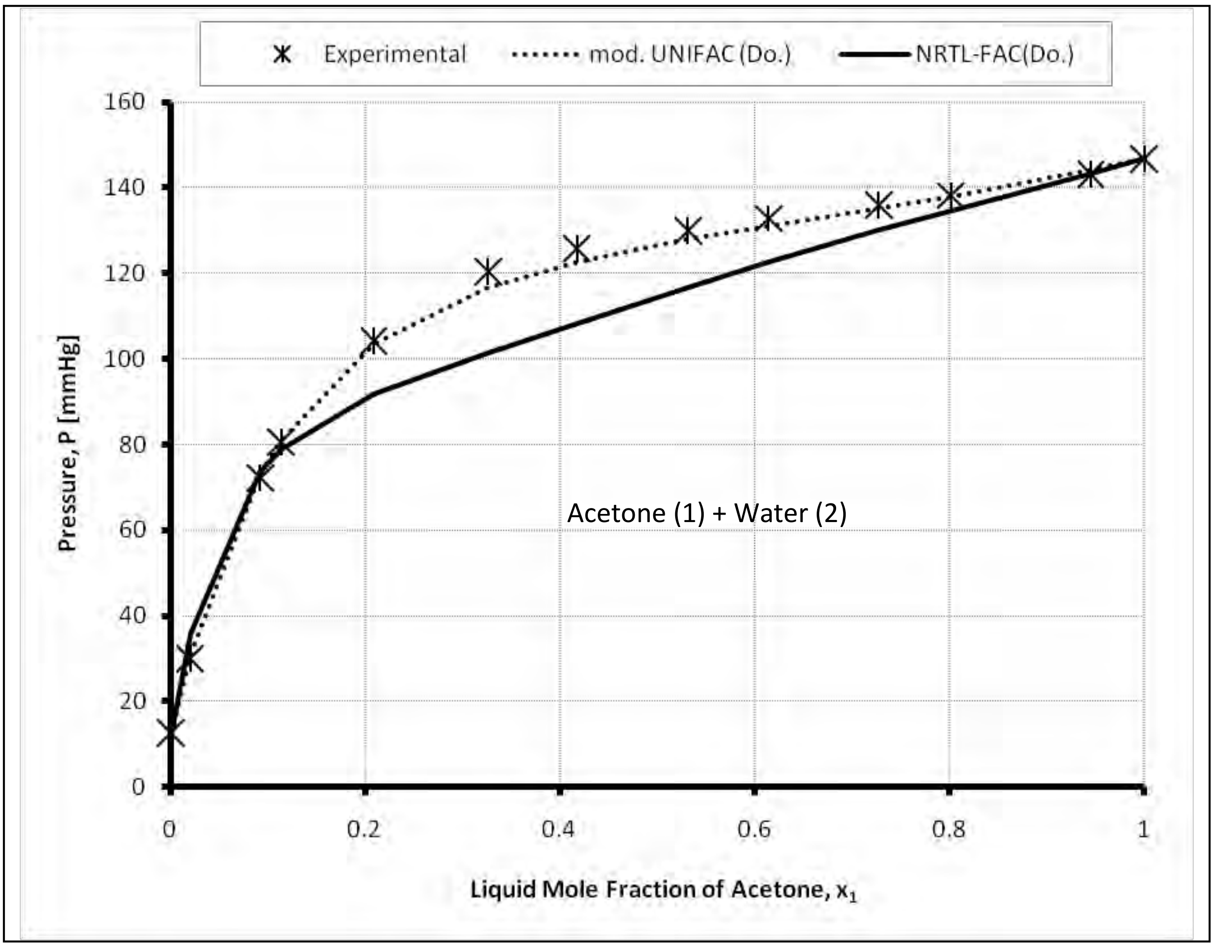


Figure 42 DDB-VLE dataset number 8224 for the system of acetone (1) + water (2) at 288.15 K. The bad regression using the NRTL-FAC(Do.) model likely has to do with the program used to fit the model parameters, where it appears that the step values used in the Simplex Nelder-Mead method had trouble reconciling the differences between the predictions of mod. UNIFAC (Do.) and the experimental data points.

Now returning to Figure 41 and Table 9, the hybrid model results are compared with their component-based alternatives. In the case of the NRTL-FAC(Do.) hybrid equation, an important improvement over the NRTL equation is seen (roughly 10%, with a crossover value of 59.3%). If this level improvement were not obtained, than the development of the hybrid models would largely have remained an academic exercise (instead of something that may be of use to practitioners). The fitting results of the Wilson-FAC(Do.) model, on the other hand, is less satisfactory (with a crossover value of 50.5%). In this case, it is basically a “toss-up” between which model will give the best representation of the experimental data. Typical examples showing the benefit of the hybrid approach using the NRTL-FAC(Do.) model are shown in Figure 43 through Figure 46 below.

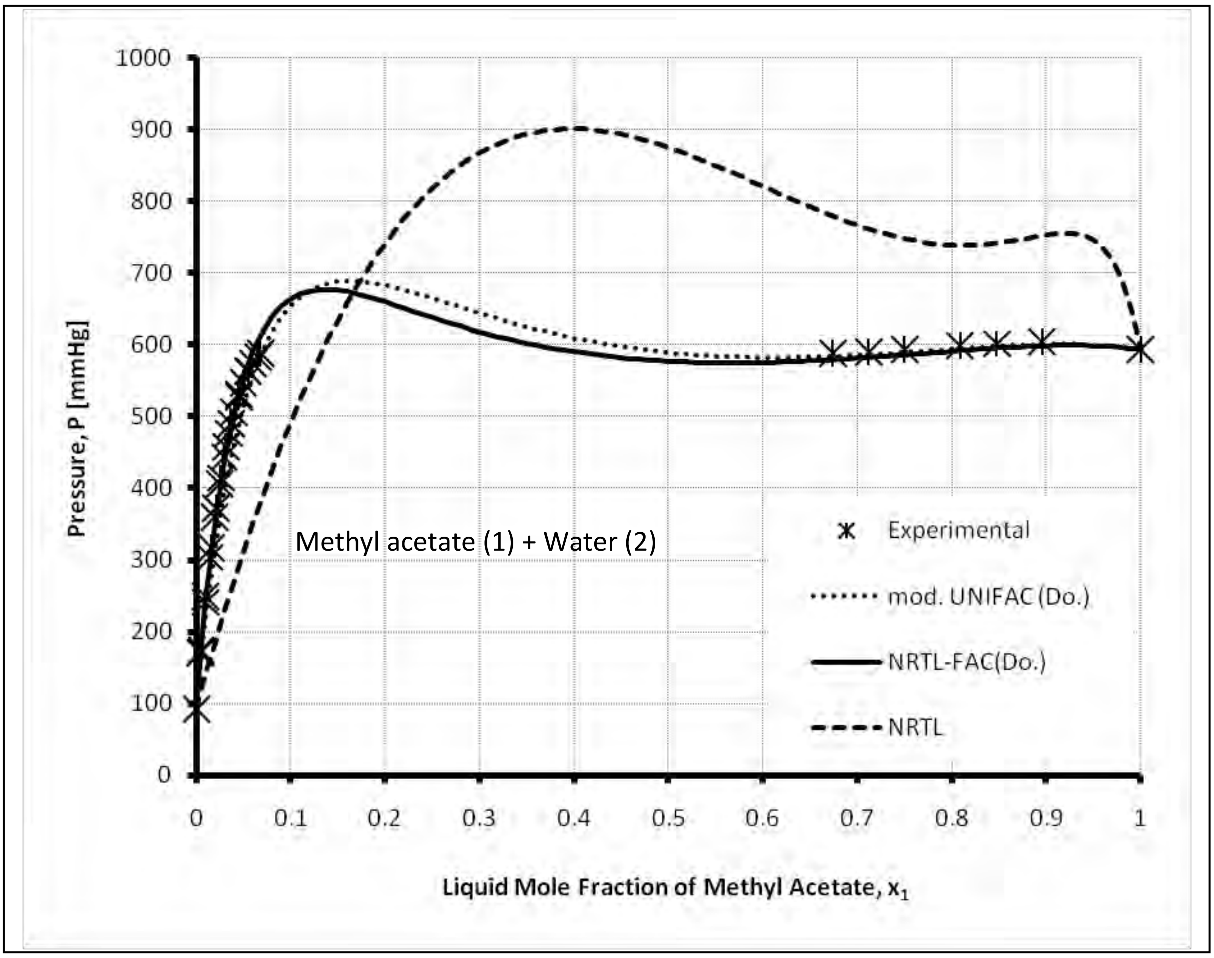


Figure 43 NRTL-FAC(Do.) hybrid model results for DDB-VLE dataset number 10763 for the system methyl acetate (1) + water (2) at 323.15 K.

As seen for the system methyl acetate (1) + water (2), while the NRTL model by itself is unable to fit to the experimental data, however very good agreement is obtained using the mod. UNIFAC (Do.) method. The hybrid model, therefore, using the NRTL model only has to make slight adjustments to the base-predictions made using the mod. UNIFAC (Do.) method. Although the NRTL model did not have to make large adjustments to the predictions, in this case, the NRTL-FAC(Do.) model still results in the best description of the experimental data (especially for lower concentrations of methyl acetate).

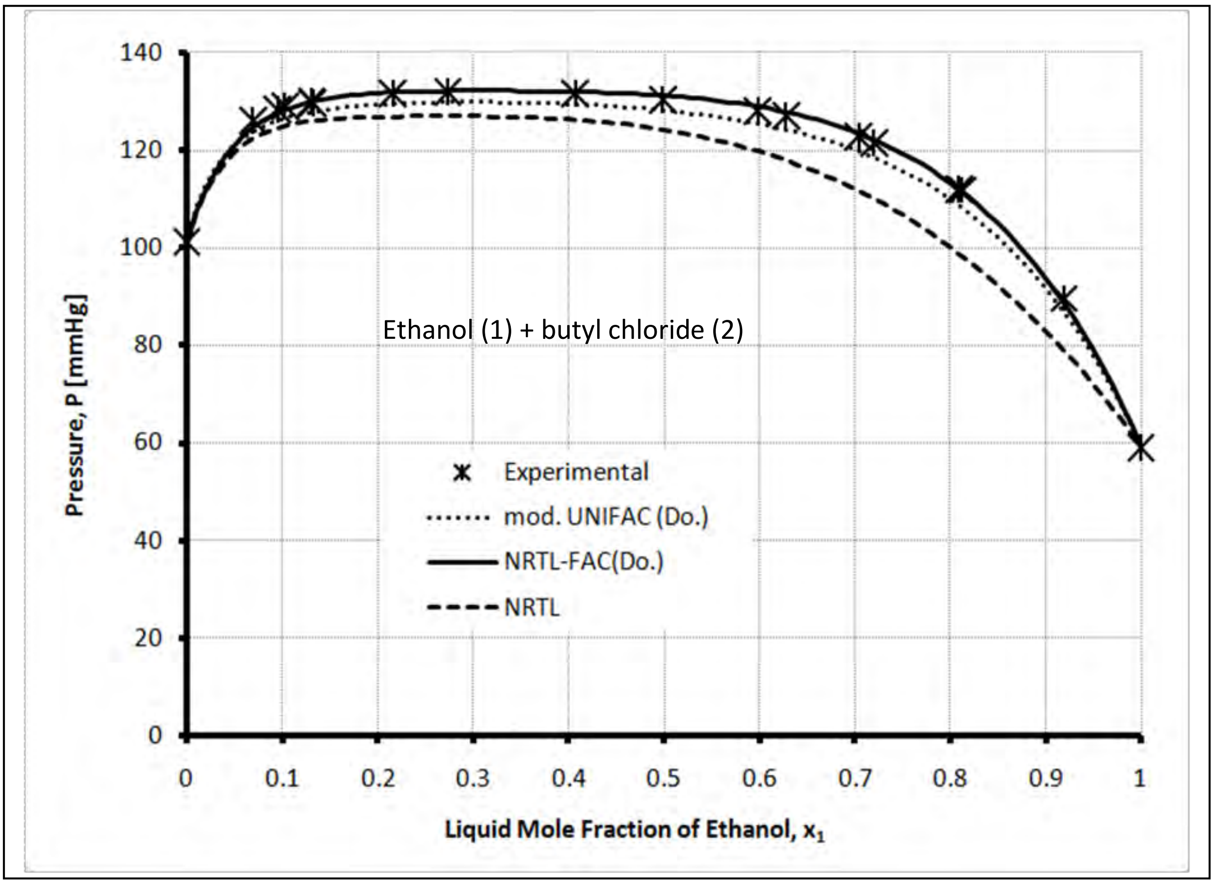


Figure 44 NRTL-FAC(Do.) hybrid model results for DDB-VLE dataset number 22041 for the system ethanol (1) +butyl chloride (2) at 298.15 K.

The system ethanol (1) + butyl chloride (2) is depicted in Figure 44, and is very similar to Figure 43. Unlike the previous system, however, the mod. UNIFAC (Do.) method does not have a near perfect fit; in this case, the NRTL will have to correct for larger differences between the base-predictions using the mod. UNIFAC (Do.) method and the available experimental data. Although the NRTL equation is unable to fit the experimental data, in this case, its combination with the predictions obtained using mod. UNIFAC (Do.) is able to achieve a very good agreement with the experimental data points.

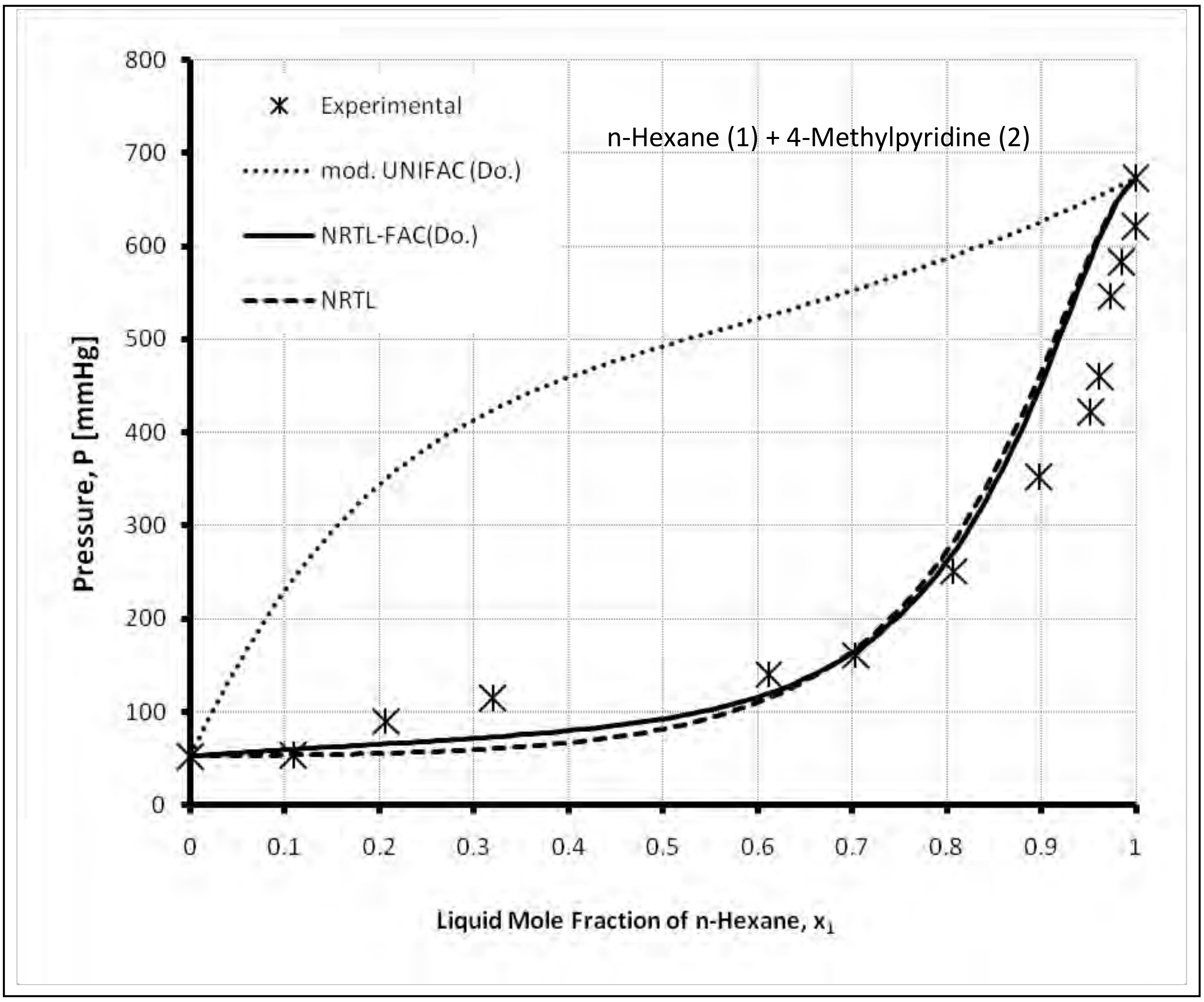


Figure 45 NRTL-FAC(Do.) hybrid model results for DDB-VLE dataset number 13199 for the system n-hexane (1) + 4-methylpyridine (2) at 338.15 K.

The system n-hexane (1) + 4-methylpyridine (2) is depicted in Figure 45 above, where the predictions made using the mod. UNIFAC (Do.) method result in the wrong description for the mixture. Not only are the errors for the predicted data points very large, but the characteristic shape for the system is completely wrong. The NRTL equation, on the other hand, obtains a fit that, although not great, is markedly better than the predictions of mod. UNIFAC (Do.). In this case the combination of these two approaches results in a fit marginally better than the fitting results of the standard NRTL equation, and remarkably better than the predictions made using the mod. UNIFAC (DO.) method. Therefore, the hybrid approach is capable of making very large adjustments to the base-predictions, so large that you can completely change the predicted characteristic shape of the curve.

An additional example showing the benefits of the hybrid approach is given in Figure 46. In this case, like Figure 43 and Figure 44, the NRTL equation is unable to obtain a good agreement with the experimental data, while the predictions made using the mod. UNIFAC (Do.) method are markedly better. These base-predictions are then further improved through the hybrid approach using the NRTL-FAC(Do.) method. Many other examples can be shown for systems in which the NRTL-FAC(Do.) method fits experimental data better than the NRTL method (59.3% of the time, see

Figure 41 (a)), but they are very similar to the figures just discussed (Figure 43 through Figure 46). Since additional figures do not add anything for discussion purposes, the few examples that have been discussed should be sufficient to prove that the base-predictions of a group contribution method, such as mod. UNIFAC (Do.), can be influenced by layering a component-based correction on top of the predictive method.

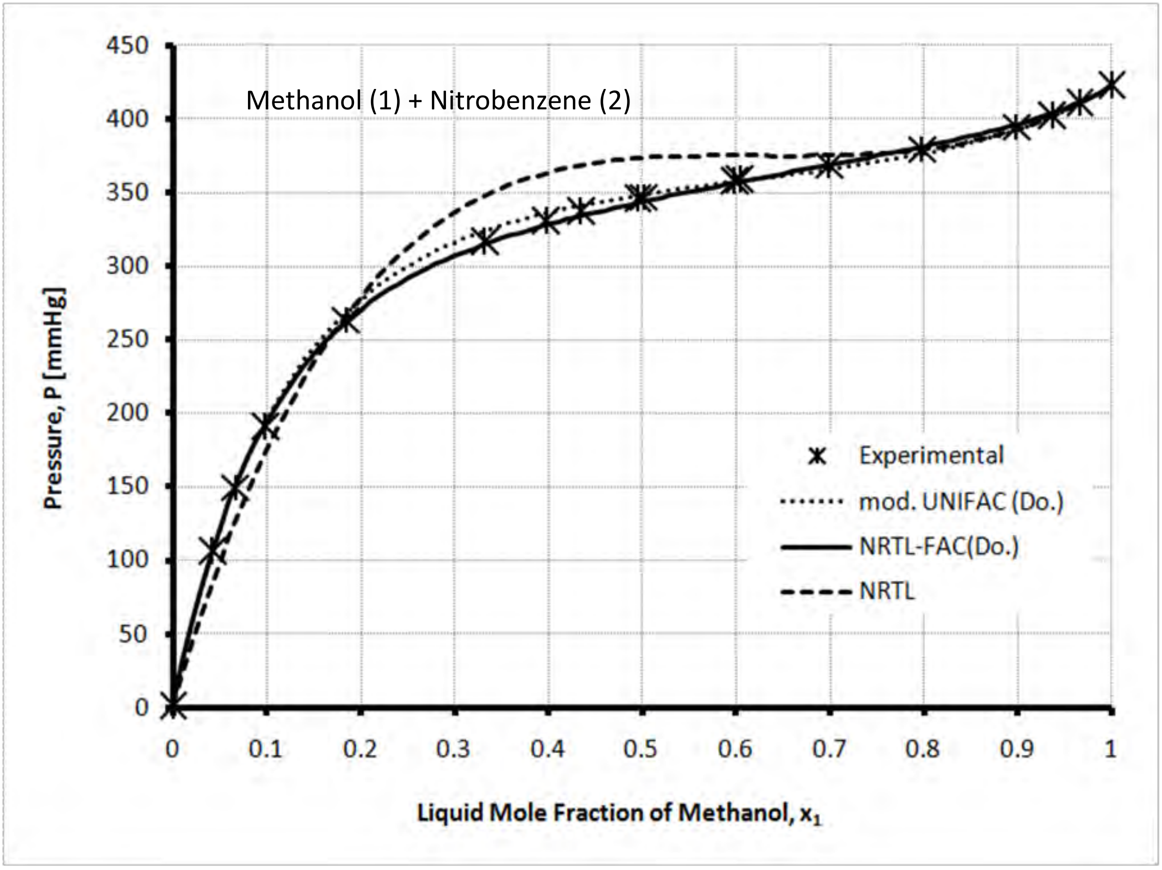


Figure 46 NRTL-FAC(Do.) hybrid model results for DDB-VLE dataset number 17187 for the system methanol (1) + nitrobenzene (2) at 323.15 K.

In some cases, however, the hybrid approach is unable to correct the predictions made using the base-predictor. In cases like these, the errors of the predicted values are too large, or the predictions result in extreme curve-shapes which cannot be adjusted to match the experimental data. Examples of this are shown in Figure 47 and Figure 48. In cases like these, it would likely be better to set the maingroup interactions to zero, so that the datasets could be regressed using *only* the component-based correction model. This would give practitioners a lot of flexibility:

This would enable practitioners to use group methods directly in their process simulations more effectively, without requiring them to perform GC2gE conversions in order to match experimental data.

In cases when the predicted errors are too large to be corrected (see Figure 47 and Figure 48), the maingroup interactions can be set to zero and the hybrid model reduces to the component-based approach.

For cases in which practitioners do not have quality experimental data available, the hybrid model reduces to the group contribution method used in the hybrid method (without the potential for a “loss in translation” occurring).

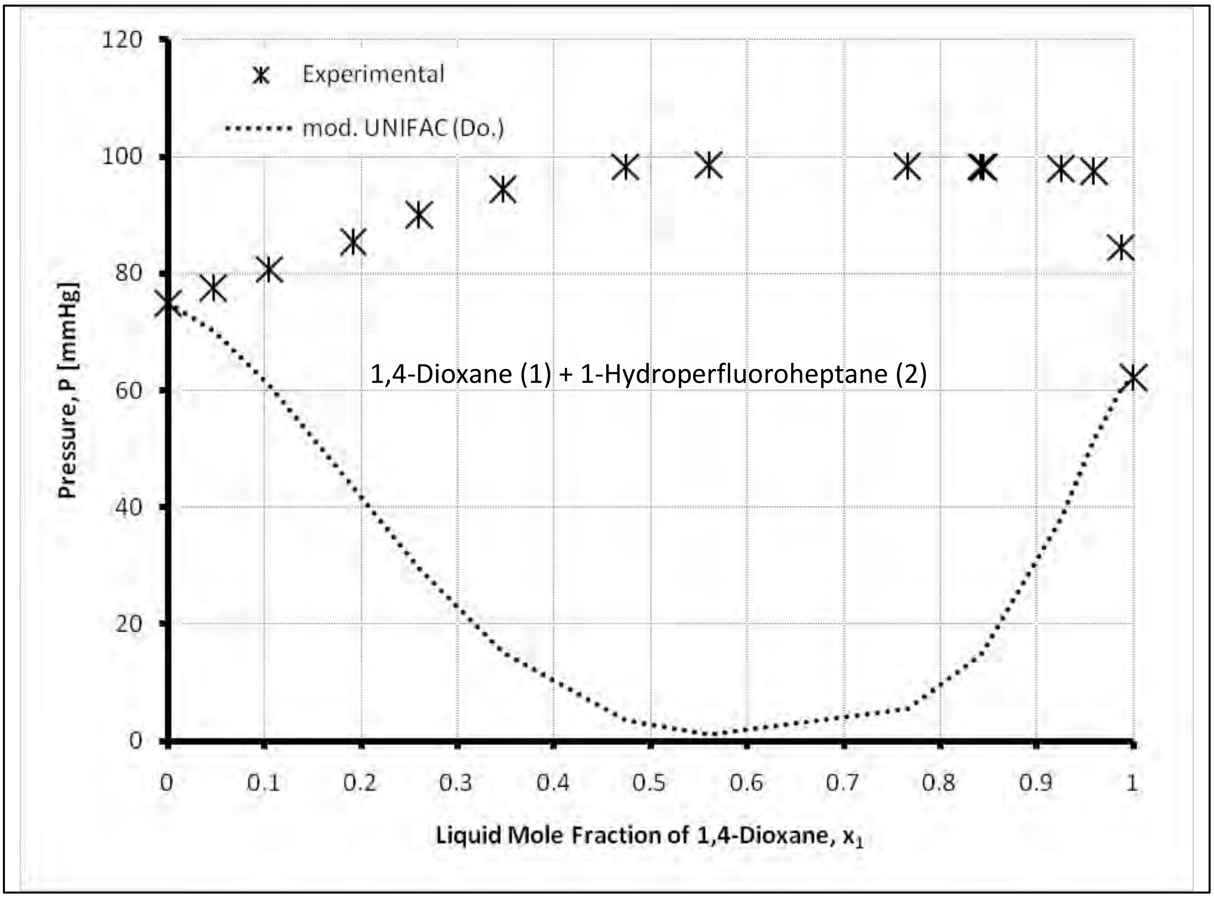


Figure 47 An illustration of predictions made using the mod. UNIFAC (Do.) method that cannot be corrected for using the NRTL-FAC(Do.) hybrid model. DDB-VLE dataset number 5557 for the system 1,4-dioxane (1) + 1-hydroperfluoroheptane (2) at 309.15 K.

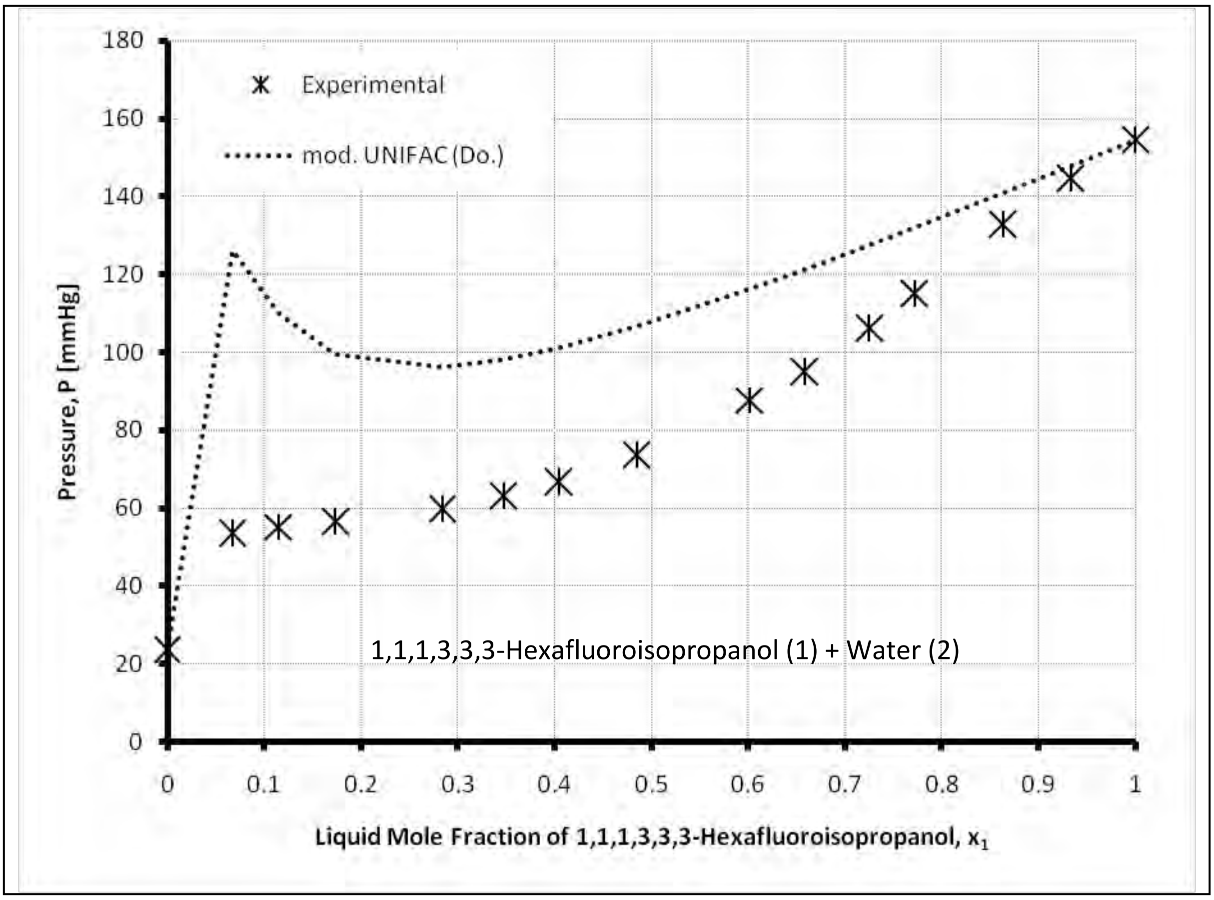


Figure 48 An illustration of predictions made using the mod. UNIFAC (Do.) method that cannot be corrected for using the NRTL-FAC(Do.) hybrid model. DDB-VLE dataset number 6483 for the system 1,1,1,3,3,3-hexafluoroisopropanol (1) + water (2) at 298.15 K.

6. Conclusions and Recommendations

6.1. Concluding Remarks

For the comparable case where a mixture can be represented by two maingroups, no benefit in modelling the system as a mixture of interacting groups was observed over modelling the system as a mixture of interacting *whole* molecules.

- For those systems defined by more than two maingroups, the benefit is seen to shift to the group-based approach. These cases represent an "apples to oranges" comparison, however, since the number of maingroup parameters being fitted will be twice as many or more compared to the component-based approaches tested.

It was seen that that the mod. UNIFAC (Do.) method provides the best representation of the experimental data that was tested, when compared against the UNIFAC method.

- All of the calculated predictions using these methods were made using the Consortium parameters (The UNIFAC Consortium, 2008).

The MRR combinatorial does not appear to be suitable as a direct combinatorial replacement for the UNIFAC or mod. UNIFAC (Do.) method, except for alkane-solvent systems.

- Based on this evaluation, it seems likely that any "tweaks" made to the combinatorial expressions of either UNIFAC or mod. UNIFAC (Do.), in order to obtain specialized/specific results, will be accompanied by a loss of generalization of these predictive methods. In most cases, the calculated combinatorial values are too coupled to the calculated residual values.

The results of the NRTL-FAC(Do.) and Wilson-FAC(Do.) hybrid models are as good as or better than each of the base models making up their parts in most cases.

6.2. Recommendations for Future Work

The results using the NRTL-FAC(Do.) and Wilson-FAC(Do.) hybrid methods seem promising, but they have only been evaluated for their ability to represent experimental data on a case-by-case basis thus far. The next step should be to incorporate the hybrid methods into a commercial process simulator such as AspenPlus, to see how they behave when used to perform typical process engineering calculations. This can be done, for instance, by the creation of a custom thermodynamic-model which can be compiled to run with AspenPlus.

The ability of these hybrid methods to represent experimental LLE data also needs to be investigated.

Given the results, it seems beneficial to include more than one component-based correction into a hybrid method (e.g. mod. UNIFAC (Do.) + NRLT + Wilson), so that practitioners can choose the component-based correction used for specific/different binary mixtures. This idea, however, has yet to be tested.

References

[1] **Abrams Denis S. and Prausnitz John M.** Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems [Journal] // AIChE Journal. - 1975. - 1 : Vol. 21. - pp. 116-128.

[2] **Alberty Robert A.** Legendre Transforms in Chemical Thermodynamics [Journal] // Pure & Appl. Chem.. - 1997. - 11 : Vol. 69. - pp. 2221-2230.

[3] **Alberty Robert A.** Use of Legendre Transformations in Chemical Thermodynamics [Journal] // Pure Appl. Chem.. - 2001. - 8 : Vol. 73. - pp. 1349-1380.

[4] **Athés Violaine [et al.]** Vapour-liquid Equilibria of Aroma Compounds in Hydroalcoholic Solutions: Measurements with a Recirculation Method and Modelling with the NRTL and COSMO-SAC Approaches [Journal] // Fluid Phase Equilibria. - 2008. - Vol. 265. - pp. 139-154.

[5] **Banerjee Sujit** Calculation of Water Solubility of Organic Compounds with UNIFAC-Derived Parameters [Journal] // Environ. Sci. Technol.. - 1985. - 4 : Vol. 19. - pp. 369-370.

[6] **Bondi A.** van der Waals Volumes and Radii [Journal] // Physical Chemistry. - 16 March 1964. - 3 : Vol. 68. - pp. 441-451. - Shell Development Company, Emeryville, California.

[7] **Bruin Solke** Activity Coefficient Relations in Miscible and Partially Miscible Multicomponent Systems [Journal] // Ind. Eng. Chem. Fundamen.. - 1970. - 3 : Vol. 9. - pp. 305-314.

[8] **Bruin Solke and Prausnitz J. M.** One-Parameter Equation for Excess Gibbs Energy of Strongly Nonideal Liquid Mixtures [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1971. - 4 : Vol. 10. - pp. 562-572.

[9] **Byer Stanley M., Gibbs Richard E. and Van Ness Hendrick C.** Vapor-Liquid Equilibrium: Part II. Correlations from P-x Data for 15 Systems [Journal] // AIChE Journal. - March 1973. - 2 : Vol. 19. - pp. 245-251.

[10] **Cadoret Loic [et al.]** Effects of Physical Properties Estimation on Process Design: A Case Study Using AspenPlus [Journal] // Asia-Pac. J. Chem. Eng.. - 2009. - Vol. 4. - pp. 729-734.

[11] **Carnot Sadi** Reflections on the Motive Power of Fire: And Other Papers on the Second Law of Thermodynamics by S. Carnot - and other Papers on the Second Law of Thermodynamics by E. Clapeyron and R. Clausius [Book] / ed. Mendoza E. / trans. Thurston R. H.. - [s.l.] : Dover Publications, 1890. - ISBN-10: 0486446417.

[12] **DDBST Software and Separation Technology GmbH** DDBST: Software and Separation Technology [Online]. - 2010. - 2009-2010. - http://www.ddbst.com/.

[13] **DDBST Software and Separation Technology GmbH** Dortmund Data Bank Software Products. - [s.l.] : DDBST GmbH, 2009.

[14] **DDBST Software and Separation Technology GmbH** Dortmund Data Bank, Version 2009 (6.1.0.360) [Software]. - [s.l.] : DDBST GmbH, 2009.

[15] **DDBST Software and Separation Technology GmbH** Prediction of Water Solubility of Hydrocarbons by modified UNIFAC (Dortmund): Improving the Prediction by Using Correction Terms [Report]. - 2009.

[16] **Deal Carl H. and Derr Elwood L.** Group Contributions in Mixtures [Journal] // Ind. Eng. Chem.. - 1968. - 4 : Vol. 60. - pp. 28-38.

[17] **Derr E. L. and Deal C. H.** Analytical Solution of Groups: Correlation of Activity Coefficients Through Structural Group Parameters [Journal] // Inst. Chem. Eng. Symp. Ser.. - 1969. - Vol. 32. - pp. 44-51.

[18] **Elliott Richard J. and Lira Carl T.** Introductory Chemical Engineering Thermodynamics [Book]. - [s.l.] : Prentice Hall PTR, 1999. - ISBN-10: 0130113867.

[19] **Falk G.** Entropy, a Resurrection of Caloric: A Look at the History of Thermodynamics [Journal] // Eur. J. Phys.. - 1985. - Vol. 6. - pp. 108-115.

[20] **Fischer Kai, Kleiber Michael and de Loos Theo W.** 60th Birthday of Jürgen Gmehling [Journal] // Fluid Phase Equilibria. - 2007. - Vol. 260. - pp. 1-2.

[21] **Fredenslund Aage [et al.]** Computerized Design of Multicomponent Distillation Columns Using the UNIFAC Group Contribution Method for Calculation of Activity Coefficients [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1977. - 4 : Vol. 16. - pp. 450-462. - DOI: 10.1021/i260064a004.

[22] **Fredenslund Aage, Gmehling Jürgen and Rasmussen Peter** Vapor-Liquid Equilibria Using UNIFAC: A Group-Contribution Method [Book]. - Amsterdam : Elsevier, 1977.

[23] **Fredenslund Aage, Jones Russell L. and Prausnitz John M.** Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures [Journal] // AIChE Journal. - 1975. - 6 : Vol. 21. - pp. 1086-1099.

[24] **Gallagher Warren** Unit 1 - Molecules: Elaboration - Molecular Polarity [Online] // Chem 150: Survey of Biochemistry - Spring 2007. - University of Wisconsin-Eau Claire, 2007. - December 2010. - http://www.chem.uwec.edu/Chem150_S07/elaborations/unit1/unit1-e-polarity.html.

[25] **Gmehling J. [et al.]** Status and Results of Group Contribution Methods [Journal] // Pure & Appl. Chem.. - 1993. - 5 : Vol. 65. - pp. 919-926.

[26] **Gmehling Jürgen [et al.]** A Modified UNIFAC (Dortmund) Model: 3. Revision and Extension [Journal] // Ind. Eng. Chem. Res.. - 1998. - 12 : Vol. 37. - pp. 4876-4882.

[27] **Gmehling Jürgen [et al.]** A Modified UNIFAC (Dortmund) Model: 4. Revision and Extension [Journal] // Ind. Eng. Chem. Res.. - 2002. - 6 : Vol. 41. - pp. 1678-1688.

[28] **Gmehling Jürgen** From UNIFAC to modified UNIFAC to PSRK with the help of DDB [Journal] // Fluid Phase Equilibria. - 1995. - Vol. 107. - pp. 1-29.

[29] **Gmehling Jürgen** Group Contribution Methods: Ideal Tools for the Synthesis and Design of Separation Processes [Journal] // Pure Appl. Chem.. - 1999. - 6 : Vol. 71. - pp. 939-949. - Lecture presented at the 7th International Chemistry Conference in Africa & 34th Convention of the South African Chemical Institute, Durban, South Africa.

[30] **Gmehling Jürgen** Present Status and Potential of Group Contribution Methods for Process Development [Journal] // J. Chem. Thermodynamics. - 2009. - Vol. 41. - pp. 731-747.

[31] **Gmehling Jürgen, Li Jiding and Schiller Martin** A Modified UNIFAC Model: 2. Present Parameter Matrix and Results for Different Thermodynamic Properties [Journal] // Ind. Eng. Chem. Res.. - 1993. - 1 : Vol. 32. - pp. 178-193.

[32] **Gmehling Jürgen, Rasmussen Peter and Fredenslund Aage** Vapor-Liquid Equilibria by UNIFAC Group Contribution: Revision and Extension no. 2 [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1982. - Vol. 21. - pp. 118-127.

[33] **Góral M.** Derivation of Wilson Equation for gE from Association Modelss [Journal] // Fluid Phase Equilibria. - 1995. - Vol. 106. - pp. 11-16.

[34] **Grensemann H. and Gmehling J.** Rebuttal to the Comments of Andreas Klamt on "Performance of a COSMO-RS Model in Comparison to Classical Group Contribution Methods" [Journal] // Ind. Eng. Chem. Res.. - 2005. - 17 : Vol. 44. - pp. 7043-7044.

[35] **Hála Eduard** Note to Bruin-Prausnitz One-Parameter and Palmer-Smith Two-Parameter Local Composition Equation [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1972. - 4 : Vol. 11. - p. 638.

[36] **Hála Eduard** Some Notes on the Group Solution Concept [Journal] // Collection Czechoslov. Chem. Commun.. - 1978. - Vol. 43. - pp. 10-15.

[37] **Hansen Henrik K. [et al.]** Vapor-Liquid Equilibria by UNIFAC Group Contribution: 5. Revision and Extension [Journal] // Ind. Eng. Chem. Res.. - 1991. - 10 : Vol. 30. - pp. 2352-2355.

[38] **Heidemann Robert A. and Mandhane Jamanlal M.** Some Properties of the NRTL Equation in Correlating Liquid-Liquid Equilibrium Data [Journal] // Chem. Eng. Sci.. - 1973. - 5 : Vol. 28. - pp. 1213-1221.

[39] **Hoheisel C. and Kohler F.** Local Composition in Liquid Mixtures [Journal] // Fluid Phase Equilibria. - 1984. - Vol. 16. - pp. 13-24.

[40] **Hu Y., Azevedo E. G. and Prausnitz J. M.** The Molecular Basis for Local Compositions in Liquid Mixture Models [Journal] // Fluid Phase Equilibria. - 1983. - Vol. 13. - pp. 351-360.

[41] **Jakob Antje [et al.]** Further Development of Modified UNIFAC (Dortmund): Revision and Extension 5 [Journal] // Ind. Eng. Chem. Res.. - 2006. - 23 : Vol. 45. - pp. 7924-7933.

[42] **Katayama Takashi, Kato Masahiro and Yasuda Minoru** Considerations on NRTL Equation as Applied to Partially Miscible Systems [Journal] // J. Chem. Eng. Japan. - 1973. - 4 : Vol. 6. - pp. 357-359.

[43] **Kemény Sándor and Rasmussen Peter** A Derivation of Local Composition Expressions from Partition Functions [Journal] // Fluid Phase Equilibria. - 1981. - Vol. 7. - pp. 197-203.

[44] **Kojima K. and Tochigi K.** Prediction of Vapor-Liquid Equilibria by the ASOG Method [Book]. - Tokyo : Kodansha Ltd., 1979. - Copublished by Elsevier Scientific, Amsterdam, The Netherlands.

[45] **Lafyatis David S. [et al.]** A Test of the Functional Dependence of gE(x) in Liquid-Liquid Equilibria Using Limiting Activity Coefficients [Journal] // Ind. Eng. Chem. Res.. - 1989. - 5 : Vol. 28. - pp. 585-590.

[46] **Lin Shiang-Tai and Sandler Stanley I.** A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model [Journal] // Ind. Eng. Chem. Res.. - 2002. - 5 : Vol. 41. - pp. 899-913.

[47] **Lohmann Jürgen, Joh Ralph and Gmehling Jürgen** From UNIFAC to Modified UNIFAC (Dortmund) [Journal] // Ind. Eng. Chem. Res.. - 2001. - 3 : Vol. 40. - pp. 957-964.

[48] **Magnussen Thomas, Rasmussen Peter and Fredenslund Aage** UNIFAC Parameter Table for Prediction of Liquid-Liquid Equilibria [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1981. - 2 : Vol. 20. - pp. 331-339.

[49] **Malanowski Stanisław and Anderko Anderzej** Modelling Phase Equilibria: Thermodynamic Background and Practical Tools [Book]. - [s.l.] : John Wiley & Sons, Inc., 1992.

[50] **Mansoori G. Ali** Statistical Mechanical Theory of Local Compositions [Journal] // Fluid Phase Equilibria. - 1985. - Vol. 22. - pp. 253-275.

[51] **Marcilla A., Gomis V. and Esteban A.** Letter to the Editor [Journal] // AIChE Journal. - April 1995. - 4 : Vol. 41. - p. 1044. - Points out that the secondary parameters, introduced by the local composition concept, is inconsistent with theory..

[52] **Marina J. M. and Tassios D. P.** Effective Local Compositions in Phase Equilibrium Correlations [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1973. - 1 : Vol. 12. - pp. 67-71.

[53] **Maurer G. and Prausnitz J. M.** On the Derivation and Extension of the UNIQUAC Equation [Journal] // Fluid Phase Equilibria. - 1978. - Vol. 2. - pp. 91-99.

[54] **McDermott C. and Ashton N.** Note on the Definition of Local Composition [Journal] // Fluid Phase Equilibria. - 1977. - pp. 33-35.

[55] **Microsoft** Microsoft Office Suite 2007. - 30 January 2007. - Service Pack 2 (12.0.6535.5002).

[56] **Moller Bruce** Activity of Complex Multifunctional Organic Compounds in Common Solvents [Report] : PhD Thesis / Chemical Engineering ; University of KwaZulu-Natal. - Durban, South Africa : [s.n.], 2010.

[57] **Mollerup Jørgen** A Note on Excess Gibbs Energy Models, Equations of State and the Local Composition Concept [Journal] // Fluid Phase Equilibria. - 1981. - Vol. 7. - pp. 121-138.

[58] **Morisue Tetsuo, Noda Katsuji and Ishida Kiyoharu** A Discussion of the Excess Gibbs Energy of Liquid Mixtures Based on the Local Composition [Journal] // J. Chem. Eng. Japan. - 1972. - 3 : Vol. 5. - pp. 219-222.

[59] **Mu Tiancheng, Rarey Jürgen and Gmehling Jürgen** Performance of COSMO-RS with Sigma Profiles from Different Model Chemistries [Journal] // Ind. Eng. Chem. Res.. - 2007. - Vol. 46. - pp. 6612-6629.

[60] **Müller Ingo** A History of Thermodynamics [Book]. - New York : Springer-Verlag Berlin Heidelberg, 2007.

[61] **Nagata Isamu and Yamada Toshiro** Predicgtion of Excess Heat Capacities of Alcohol-Hydrocarbon Mixtures [Journal] // J. Chem. Eng. Data. - 1973. - 1 : Vol. 18. - pp. 87-90.

[62] **Naidoo Thishendren** Application of a non-Linear Transformation to the Surface Fraction of the UNIQUAC Model and the Performance Analysis of the Subsequent Model (FlexQUAC-Q) [Report] : MSc Thesis / Chemical Engineering ; University of Kwazulu-Natal. - Durban : [s.n.], 2007.

[63] **NASA** Researchers Cool Gas to Record Low [Online] // NASA. - 11 September 2003. - 22 Febuary 2009. - http://www.nasa.gov/vision/earth/technologies/biggest_chill.html.

[64] **Nelder J. A. and Mead R.** A Simplex Method for Function Minimization [Journal] // The Computer Journal. - 1965. - 4 : Vol. 7. - pp. 308-313. - doi: 10.1093/comjnl/7.4.308.

[65] **Orye R. V. and Prausnitz J. M.** Multicomponent Equilibria with the Wilson Equation [Journal] // Ind. & Eng. Chem.. - 1965. - 5 : Vol. 57. - pp. 18-26.

[66] **Poling Bruce E., Prausnitz John M. and O'Connell John P.** The Properties of Gases and Liquids [Book]. - [s.l.] : McGraw-Hill, 2001. - 5th Edition. - ISBN 0-07-011682-2.

[67] **Porter A. W.** The Vapour-Pressures of Mixtures [Journal] // Trans. Faraday Soc.. - 1921. - Vol. 16. - pp. 336-345.

[68] **Prausnitz John M., Lichtenthaler Ruediger N. and de Azevedo Edmundo Gomes** Molecular Thermodynamics of Fluid-Phase Equilibria [Book]. - Englewood Cliffs : Prentice-Hall Inc., 1986. - 2nd Edition.

[69] **Rarey Jürgen** Extended Flexibility for gE Models and Simultaneous Description of Vapor-Liquid Equilibrium and Liquid-Liquid Equilibrium Using a Nonlinear Transformation of the Concentration Dependence [Journal] // Ind. Eng. Chem. Res.. - 2005. - 19 : Vol. 44. - pp. 7600-7608.

[70] **Rarey Jürgen** Fundamentals of Selection and Design of Thermal Separation Processes and DDB/DDBSP Course [Presentation/Course]. - Galveston : [s.n.], 7-8 December 2006. - Sponsered by EPCON International.

[71] **Rarey Jürgen** Personal Communications. - 2009-2010.

[72] **Redlich Otto and Kister A. T.** Algebraic Representation of Thermodynamic Properties and the Classification of Solutions [Journal] // Ind. & Eng. Chem.. - February 1948. - 2 : Vol. 40. - pp. 345-348.

[73] **Renon Henri and Prausnitz J. M.** Estimation of Parameters for the NRTL Equation for Excess Gibbs Energies of Strongly Nonideal Liquid Mixtures [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1969. - 3 : Vol. 8. - pp. 413-419.

[74] **Renon Henri and Prausnitz J. M.** Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures [Journal] // AIChE Journal. - 1968. - 1 : Vol. 14. - pp. 135-144.

[75] **Rizzi Andreas and Huber Josef F. K.** Comparative Calculations of Activity Coefficients in Binary Liquid Mixtures at Infinite Dilution Using the "Solution of Groups" Model [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1981. - Vol. 20. - pp. 204-210.

[76] **Ronc M. and Ratcliff G. A.** Prediction of Excess Free Energies of Liquid Mixtures by an Analytical Group Solution Model [Journal] // Can. J. Chem. Eng.. - 1971. - 6 : Vol. 49. - pp. 825-830.

[77] **Scatchard George and Wilson Grant** Vapor-Liquid Equilibrium: XIII. The System Water-Butyl Glycol from 5 to 85 Degrees [Journal] // J. Am. Chem. Soc.. - 1964. - 2 : Vol. 86. - pp. 133-137.

[78] **Scott Robert L.** Corresponding States Treatment of Nonelectrolyte Solutions [Journal] // J. Chem. Phys.. - 1956. - 2 : Vol. 25. - pp. 193-205.

[79] **Sherman Steven R. [et al.]** Compilation and Correlation of Limiting Activity Coefficients of Nonelectrolytes in Water [Journal] // Ind. Eng. Chem. Res.. - 1996. - Vol. 35. - pp. 1044-1058.

[80] **Srinivasan J.** Sadi Carnot and the Second Law of Thermodynamics [Journal] // Resonance. - November 2001. - pp. 42-48.

[81] **Tassios Dimitrios** Limitaitons in Correlating Strongly Nonideal Binary Systems with the NRTL and LEMF Equations [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1976. - 4 : Vol. 15. - pp. 574-578.

[82] **The UNIFAC Consortium** 2008 UNIFAC and mod. UNIFAC (Do) Consortium Parameters. - [s.l.] : University of Oldenburg Department of Industrial Chemistry, 2008.

[83] **Thomas Eugene R. and Eckert Charles A.** Prediciton of Limiting Activity Coefficients by a Modified Separation of Cohesive Energy Density Model and UNIFAC [Journal] // Ind. Eng. Chem. Process Des. Dev.. - 1984. - Vol. 23. - pp. 194-209.

[84] **Tochigi K. [et al.]** The Temperature Dependence of ASOG Parameters VLE Calculations [Journal] // AIChE Journal. - 1981. - 6 : Vol. 27. - pp. 1022-1024.

[85] **Tochigi Katsumi and Kojima Kazuo** The Determination of Group Wilson Parameters to Activity Coefficients by Ebulliometer [Journal] // J. Chem. Eng. Japan. - 1976. - 4 : Vol. 9. - pp. 267-273.

[86] **Tochigi Katsumi, Minami Shigeki and Kojima Kazuo** Prediciton of Vapor-Liquid Equiibria with Chemical Reaction by Analytical Solutions of Groups [Journal] // J. Chem. Eng. Japan. - 1977. - 5 : Vol. 10. - pp. 349-354.

[87] **Van Ness Hendrick C. and Abbott Michael M.** Classical Thermodynamics of Nonelectrolyte Solutions: With Applications to Phase Equilibria [Book]. - [s.l.] : McGraw-Hill, 1982. - McGraw-Hill Chemical Engineering Series. - ISBN 0-07-067095-1.

[88] **Van Ness Hendrick C.** Thermodynamics in the Treatment of Vapor/Liquid Equilibrium (VLE) Data [Journal] // Pure & Appl. Chem.. - 1995. - 6 : Vol. 67. - pp. 859-872.

[89] **Walas Stanley M.** Phase Equilibria in Chemical Engineering [Book]. - [s.l.] : Butterworth-Heinemann, 1985.

[90] **Weidlich Ulrich and Gmehling Jürgen** A Modified UNIFAC Model. 1. Prediction of VLE, hE, and Activity Coefficients at Infinite Dilution [Journal] // Ind. Eng. Chem. Res.. - 1987. - Vol. 26. - pp. 1372-1381.

[91] **Wilson G. M. and Deal C. H.** Activity Coefficients and Molecular Structure: Activity Coefficients in Changing Environments - Solution of Groups [Journal] // Ind. Eng. Chem. Fundam.. - 1962. - 1 : Vol. 1. - pp. 20-23.

[92] **Wilson Grant M.** Vapor-Liquid Equilibrium: XI. A New Expression for the Excess Free Energy of Mixing [Journal] // J. Am. Chem. Soc.. - 1964. - Vol. 86. - pp. 127-130.

[93] **Wittig Roland, Lohmann Jürgen and Gmehling Jürgen** Vapor-Liquid Equilibria by UNIFAC Group Contribution. 6. Revision and Extension [Article] // Ind. Eng. Chem. Res.. - 2003. - 1 : Vol. 42. - pp. 183-188.

[94] **Wu Huey S. and Sandler Stanley I.** Use of ab Initio Quantum Mechanics Calculations in Group Contribution Methods: 1. Theory and the Basis for Group Identificaitons [Journal] // Ind. Eng. Chem. Res.. - 1991. - Vol. 30. - pp. 881-889.

[95] **Wu Huey S. and Sandler Stanley I.** Use of ab Initio Quantum Mechanics Calculations in Group Contribution Methods: 2. Test of New Groups in UNIFAC [Journal] // Ind. Eng. Chem. Res.. - 1991. - Vol. 30. - pp. 889-897.