

A Computer Aided Framework for Prediction of Properties of Organic Systems

Hugo E. González Villalba^a, Jens Abildskov^a, Rafiqul Gani^{a*}

^aCAPEC, Department of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark, *rag@kt.dtu.dk

Abstract

This paper presents a systematic computer aided framework for modeling and prediction of pure component and mixture properties for a very wide range of chemicals and properties. Together with the extended GC⁺ property models for pure component properties and mixtures, the framework provides a feature for prediction of missing group interaction parameters (for mixture properties) by employing an atom-connectivity approach. The framework also provides a model parameter regression feature to fine-tune existing parameters and regress new parameters.

Keywords

Molecular descriptors, Connectivity Indices, UNIFAC, GC⁺

1. Introduction

Properties of chemicals play a very important role in the design of chemical product and in the design of chemical processes that can manufacture them. In computer-aided model-based methods and tools for synthesis and design of the product-process, the applicability of the model and its reliability is governed by the property models used in the product-process model. In addition to reliability of the property models, another important need in product-process synthesis and design is that the models need to be predictive and fairly simple and easy to use. GC (group contribution) models have the ability to handle a large number of

organic systems (pure compounds and mixtures). For prediction of properties of high-value chemicals encountered in food, pharmaceutical etc, industries, however, there are some limitations of GC models.

2. Problem Statement, background

Group contribution (GC) methods are based on the additivity principle of the molecular fragments (the groups) constituting the compounds. The simplicity and versatility of GC-methods have increased its popularity and use in the field of pure component and mixtures property prediction. Computational time is low and the prediction accuracies are generally acceptable. However, there are some limitations with respect to the objectives of wide application and prediction for properties involving new and complex compounds: for example, GC methods are unable to consider 3D structural parameters of molecules, and therefore, cannot predict the steric effects of isomers, which may be important in their recovery as a product.

3. GC-Plus approach

3.1. Methodology

The methodology for the construction of a GC⁺ model involves the characterization of the molecular descriptor that serves as a corrective-complementary term for the improvement on the original GC-model prediction. The methodology includes the selection of molecular descriptors related to specific characteristics of the property under study. A compromise between accuracy and engineering utility is needed. So it is necessary to develop a model that can be easily integrated to other process and product models.

3.2. Development of GC-Plus model for cis-trans pure component prediction.

Cis-trans pure component prediction.

Currently, cis/trans isomer distinction can not be made with groups alone, because groups do not describe different spatial conformations. To overcome this, relationships between structural properties and thermodynamic properties for cis/trans isomers have been explored. Using this knowledge we are developing additional correction terms to add to an existing GC method, to take into account 3D of cis/trans isomers. The work includes the: 1) Development of a new database for cis/trans isomers including atom interactions, steric properties and intermolecular forces, generated through a set of molecular level calculations and related with specific properties of interest; 2) Derivation of a

new term for addition to the original GC-model and 3) Implementation of this term into a GC⁺ method. In this way, using a large database, distinction between cis/trans isomer properties is made with good accuracy and easy-to-use software.

The Marrero-Gani^[1] (MG) method has been selected as a base GC-method and the property selected for this case study is the normal boiling point (T_b):

$$\exp \frac{T_b}{t_{b0}} = \sum_{i=1}^{1.order\ groups} N_i C_i + \sum_{i=1}^{2.order\ groups} M_i D_i + \sum_{i=1}^{3.order\ groups} L_i E_i \quad (1)$$

For the evaluation of molecular descriptors we have selected the MM2 force field (from CHEM 3D[®]) as the complementary tool for the MG model. The MM2 force field method takes into account molecular features such as the ability of bonds to stretch, bend or twist. The force field also takes into account the possibility of non-bonded interactions such as repulsion and attraction forces. So the energy associated to the force field can be described by the equation: Energy = stretching + bending + torsion + non-bonded.

The proposed model for the estimation of T_b using these molecular parameters is the following:

$$\exp \left(\frac{T_b}{t_{b0}} \right) = \exp \left(\frac{T_b}{t_{b0}} \right)_{Marrero/Gani} + \sum_{i=1} a_i N_i^A w b_i + b \cdot w b_i + c \cdot bend_i + d \cdot torsion_i \quad (2)$$

Here

wb=vdW+non vdW (van der Waals forces + non-van der Waals forces)

a...d : regressed correlation coefficients

i: compound-i

A: Atom – A

N_i^A : Atoms-of-type-A-in-molecule-i

$\exp(T_b/t_{b0})_{Marrero/Gani}$ denotes the estimate with the Marrero/Gani model on present form.

3.2.1. Results and discussion.

The results from the regression of the model parameters have been analyzed quantitatively as well as qualitatively. Based on the sum of squares error (SSE), calculated in terms of correlated value and the corresponding experimental value, the SSE using the MG-method is 8181.75 while with MG^{Plus} is reduced to 7172.53.

In the qualitative analysis, the trend of T_b as a function of the cis-trans isomer has been verified. In nearly 60 % of the isomers, the trend is correlated correctly. For the remaining data sets, the differences in the cis-trans isomer properties are not large enough to make them significant.

3.3. UNIFAC-CI Group Contribution Method

Regarding GC^+ models for mixtures; González-Villalba et.al.^[2] proposed the use of connectivity indices (CI) to overcome the lack of group interaction parameters (GIP's) for the UNIFAC GC^+ -method for VLE calculations. The basic idea behind this work is to derive a relation between the group interaction parameters with the CI and atom constitution of the groups describing the molecules under study. In this way, with a database of CI and the number of atoms, the missing GIP's can be predicted through a derived relation (see eq. 3), which we call the atom interaction function (AIF).

Once an organic chemical system is defined; for example, a system for VLE calculations involving C,O,H and corresponding group information (in terms of CI and number of atoms) and the regressed AIF , it is possible to: 1) Generate values for missing GIP's on the UNIFAC matrix 2) Reestimate one or more GIP's 3) Create a group and estimate its GIP's. González-Villalba et.al.^[2] have already generated the AIF for the UNIFAC-VLE groups involving C,O,H atoms. In this work, nitrogen related compounds have been added to the UNIFAC model. At this moment the following UNIFAC groups have been covered: CH₂, C=C, ACH, ACCH₂, OH, CH₃OH, H₂O, ACOH, CH₂CO, COH, CCOO, HCOO, CH₂O, CH₂O, CNH₂, (C)₂NH, (C)₃N, ACNH₂, CCN, and CNO₂. All these groups involves 460 GIP's. However, with the UNIFAC-CI approach, only a set of 72 AIP's need to be regressed to cover the same groups. For two groups represented by k and l, the following relation apply:

$$\begin{aligned}
 a_{kl} = & \underbrace{b_{c-c}(A_{kl}^{cc})_0 + b_{c-o}(A_{kl}^{co})_0 + b_{c-n}(A_{kl}^{cn})_0 + \dots}_{0^{\text{th}}\text{-order interactions}} + \underbrace{c_{c-c}(A_{kl}^{cc})_1 + c_{c-o}(A_{kl}^{co})_1 + c_{c-n}(A_{kl}^{cn})_1 + \dots}_{1^{\text{st}}\text{-order interactions}} \\
 & + \underbrace{d_{c-c}(A_{kl}^{cc})_2 + d_{c-o}(A_{kl}^{co})_2 + d_{c-n}(A_{kl}^{cn})_2 + \dots}_{2^{\text{nd}}\text{-order interactions}} + \underbrace{e_{c-c}(A_{kl}^{cc})_3 + e_{c-o}(A_{kl}^{co})_3 + e_{c-n}(A_{kl}^{cn})_3 + \dots}_{3^{\text{rd}}\text{-order interactions}} + \dots
 \end{aligned} \quad (3)$$

As shown in Eq. (3), the atom interaction functions, require values of atom interaction parameters (AIPs), which are obtained through:

$$(A_{kl}^{XY})_0 = \frac{n_X^{(k) \vee} \chi_{(l)}^0 - n_Y^{(l) \vee} \chi_{(k)}^0}{\vee \chi_{(l)}^0 \vee \chi_{(k)}^0} \quad (4)$$

Where the AIPs for atoms X and Y are calculated. Here $n_X^{(k)}$ is the number of atoms of type X in the group k, and $\vee X_{(x)}^m$ is the m-th order valence connectivity index for the group k.

3.3.1. Parameter optimization

From the databases available to CAPEC, 32 binary datasets of data (462 values) for VLE systems involving C,O,H,N atoms and 227 binary measurement series (3946 values) for VLE systems involving C,O,H atoms have been selected for regression of the AIP's. The parameter estimation of this work is exclusively based on isothermal data and checked for thermodynamic consistency (Van Ness^[3]). Details of the parameter optimization is given by González-Villalba et.al. ^[2].

3.3.2. Results and discussion.

Figure 1 presents the comparison between the deviation from experimental data with the original UNIFAC method, and UNIFAC-CI, for each of the data set used in the parameter estimation process. The deviation plotted is the average absolute relative deviation (AARD %) for the data set. The range of deviation using the AIP's is comparable with the original UNIFAC method, also must be pointed out that the UNIFAC-CI method does not require any extra experimental data than the one that has been extracted from published literature. Figure 2 includes results from the original UNIFAC method and UNIFAC-CI for VLE of n-butyronitrile/2-butanol^[4] system at 298.15 K. UNIFAC-CI is clearly more accurate and represents in a good way the composition near to the azeotrope.

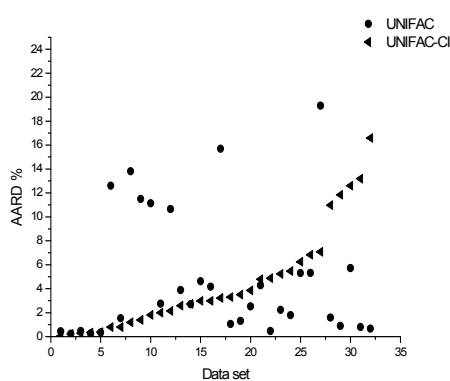


Figure 1. AARD % for both original UNIFAC AND UNIFAC-CI.

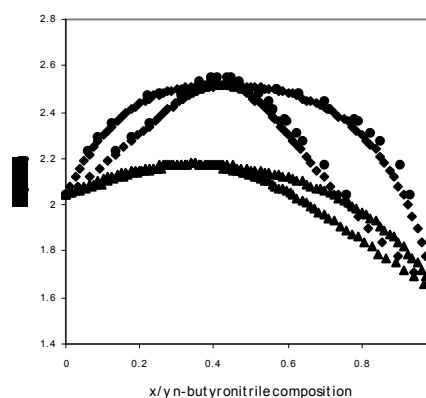


Figure 2. Comparison of the original UNIFAC method (\blacktriangle) and UNIFAC-CI (\blacklozenge) towards a n-butyronitrile/2-butanol experimental data set (\bullet).

4. Conclusions and future work

As basis for a framework for a computer-aided property prediction package, GC⁺ models have been developed and tested for the improvement on property estimations for pure compounds and mixtures. In addition, the models have been compared with well known GC models and found to provide the desired improvements. The next logical step is the further refinement of the method to improve the accuracy of the models, the extension to other properties and the implementation of the package as a computer aided tool into ICAS (www.capec.kt.dtu.dk), so that they can be used in chemical product-process design. This is particularly important since the wider application range will allow a wider search space for the synthesis/design problems.

For the T_b GC⁺ model, it is necessary to create a larger database in order to increase the reliability of the calculations. For the UNIFAC-CI model, further work includes the extension to groups including atoms other than C,O,H,N, for example halides. This is dependent on the amount of available data. Also, the extension of this method to models for LLE and SLE are being considered, together with an automated fashion for create and fit GIP's using AIP's.

Acknowledgements

Funding for Hugo E. González Villalba through the National Council of Science and Technology (CONACyT) from Mexico is gratefully acknowledged.

References

1. Marrero, J.; Gani, R. A Group Contribution Based Estimation of Pure Component Properties. *Fluid Phase Equilibria*. 2001, 183, 183.
2. González-Villalba H., Le Bert B., Gani R., Abildskov J., Prediction of Missing UNIFAC Group Interaction Parameters through Connectivity Indices, *AIChE J.* (submitted) 2006.
3. Van Ness, H.C. Thermodynamics in the treatment of (vapor + liquid) equilibria. *J. Chem. Thermodynamics*. 1995, 27, 113-134.
4. Garriga R., Sánchez F., Pérez P., Gracia M., *J. Chem. Eng. Data*, 1997, 42, 1,80.