Phase Equilibrium Calculation of Bio-oil Related Molecules Using Predictive Thermodynamic Models

Amir Jalalinejad[1†], *Jaber Yousefi Seyf*[2‡], *Axel Funke*[1†], *Nicolaus Dahmen*[1†]

[†]Institute of Catalysis Research and Technology (IKFT), Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

[‡]Department of Chemical Engineering, Hamedan University of Technology, P.O. Box: 65155-579, Hamedan, Iran

ABSTRACT: In the present study predictive thermodynamic models including original UNIFAC, Dortmund-modified UNIFAC (UNIFAC-DMD), NIST-modified UNIFAC (NIST-UNIFAC), and COSMO segment activity coefficient (COSMA-SAC) were used to predict the phase equilibrium of binary and ternary mixtures relevant for the description of fast pyrolysis bio-oils. A total of 3371, binary vapor-liquid equilibrium (VLE) isothermal or isobaric data sets were used to study the predictive power of the investigated models. Based on the obtained deviation for VLE of binary mixtures, the NIST-UNIFAC is recommended for class I mixtures (including bio-oil and non-bio-oil molecules), while the COSMO-SAC model is suggested for class II mixtures (bio-oil molecules). In sequence, 62 available ternary vapor-liquid equilibrium (VLE) isobaric data including bio-oil related and non-bio-oil molecules were used to investigate the models. Results showed that both the UNIFAC-DMD and NIST-UNIFAC provide the lowest deviation compared to UNIFAC and COSMO-SAC. Also, the COSMO-SAC model requires a large CPU time (51 minutes) for 62 ternary systems, while the NIST-UNIFAC with a CPU time of 13.2 seconds is the fastest activity coefficient model. To further compare the models ability, 125 binary LLE systems with 850 binary data sets and 2543 data points were gathered from the literature. In terms of CPU time, the group contribution models are of the order of seconds, but the COSMO-SAC model is of the order of hours (1.91 hrs). Also, the NIST-UNIFAC model provides the lowest deviation (with a slight difference from the UNIFAC-DMD model) compared to other models. In many cases, the COSMO-SAC model cannot predict the phase split of binary LLE data. Finally, 157 ternary combinations with 276 experimental ternary LLE data were collected and the results show that the UNIFAC-DMD and NIST-UNIFAC are the models with the lowest deviation. In summary, according to the obtained results in VLE systems for the bio-oil-related molecules with polar and complex structures, the group contribution

models should be used with more investigation, and the COSMO-SAC model is not recommended for LLE systems at all.

KEYWORDS: *phase equilibrium, VLE, LLE, Bio-oil, UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC*

1. INTRODUCTION

The current global landscape in terms of energy and chemical consumption presents humanity with significant challenges. Two key issues are the depleting reserves of fossil fuels and the ongoing global warming. To address these challenges, there is a growing commitment to achieve carbon neutrality by 2050, which underscores the need to shift away from relying on fossil fuels and move towards an economy centered on renewable resources. This shift is further emphasized by the fact that global energy consumption has increased notably over the past decade, in contrast to the declining availability of fossil fuel reserves. One promising approach that has emerged over the past few decades is the concept of a biorefinery.[1] There are many biorefinery concepts, and one can be based on fast pyrolysis, where biomass is efficiently converted into bio-oil.[2-4] During fast pyrolysis, biomass undergoes non-selective, thermochemical reactions, leading to the formation of a wide range of organic chemical species. The condensation of these vapors ultimately gives rise to a liquid product commonly referred to as bio-oil or pyrolysis-oil. This bio-oil is a complex mixture, featuring a multitude of oxygenated species such as water, alcohols, sugars, acids, aldehydes, ketones, pyrans, low-molecular-weight (LMM) lignin, and high-molecular-weight (HMM) lignin derivatives.[4, 5] To improve the quality of the bio-oil, e.g. by reducing oxygen content and reactivity, catalytic upgrading[6] is sometimes employed. This upgrading process can take place either *in situ* within the fast pyrolysis reactor or *ex situ* in a separate reactor downstream of the fast pyrolysis reactor.[7, 8]

The current use of fast pyrolysis bio-oil (FPBO) is primarily as fuel for industrial-scale boilers. It is also possible to co-process FPBO in existing oil refineries, directly increasing the share of renewable carbon in their product range of fuels and chemicals.[9] Another alternative is to use FPBO as fuel for a centralized gasifier to produce synthetic fuels e.g. via Fischer-Tropsch synthesis.[10-13] In order to target higher-value applications, the separation and purification of bio-oil components become pivotal[14] However, separating chemicals from such a complex mixture,

especially when they exist at low concentrations, presents formidable challenges.[15, 16] Consequently, the effective design of separation processes, necessitates the availability of thermodynamic data, particularly data related to phase equilibria.[17] Because of the high cost of the experimental work and also the nature of bio-oil-related molecules (molecules with high melting points for vapor-liquid equilibrium study), the availability of experimental phase equilibria data is limited. Therefore, thermodynamic models are an alternative to experimental research to reduce costs and save time.[18, 19] Various Gibbs free energy models have been developed for phase equilibria calculations of polar and complex mixtures, including correlative, predictive, and pure predictive models. Notably, the Wilson,[20] NRTL,[21] and UNIQUAC[22] models are well-established correlative Local Composition models. These models, relying on the specific components within a mixture, require experimental data to fine-tune their binary interaction parameters. Local Composition models are recommended when experimental data is available, but for n number of components, they require $n\times(n-1)/2$ binary interaction parameters and also the experimental pair data. In lack of experimental data, predictive models such as Analytical Solutions of Groups (ASOG), UNIQUAC Functional Group Activity Coefficients (UNIFAC),[23] UNIFAC-DMD,[24-29] NIST-modified UNIFAC,[30] and NRTL functional activity coefficient (NRTL-FAC)[31, 32] are used. Although these models are called predictive, the group-group interaction parameters, first, are fitted to many experimental data. Besides, the quality of the used experimental data is very important for both local composition and predictive models. So, when there is a limited set of experimental data, the results of these models must be used with caution. In this case, purely predictive models based on quantum chemical principles, like Conductor like Screening Model for Real Solvents (COSMO-RS)[33] and COSMO-segment activity coefficient (SAC)[34] serve as viable alternatives.

The present study aims to evaluate the predictive power of the original UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for calculation phase equilibria of bio-oil related components using a large set of binary and ternary vapor-liquid equilibrium (VLE) and binary and ternary liquid-liquid equilibrium (LLE) experimental data. Guidance for users to choose adequate phase equilibrium models is provided in consequence, which is the basis for designing unit operations in biorefinery plants. At the same time, important shortcomings are outlined to further advance the field.

2. METHOD

2.1. Models

In the present research the original UNIFAC[23], UNIFAC-DMD,[15] and the NIST-UNIFAC[30] models are applied as predictive activity coefficient models to investigate their prediction power. Although the details of these models can be found in the literature, a summary of these models will be given here. The activity coefficient of component i (γ_i) in a mixture is divided into two parts, including combinatorial and residual. The former is due to the differences in shape and size of the molecules in the mixture, while the latter is due to the differences in intermolecular interactions relative to their pure components. So, γ_i is expressed as:

$$\gamma_i = \gamma_i^C \gamma_i^{res} \tag{1}$$

Here γ_i^C, and γ_i^{res} stand for the combinatorial and residual parts of the activity coefficient, respectively. The combinatorial part of the original UNIFAC[23] is:

$$\ln \gamma_i^C = 1 - \frac{\Phi_i}{x_i} + \ln\left(\frac{\Phi_i}{x_i}\right) - \frac{Z}{2}\left[1 - \frac{\Phi_i}{\theta_i} + \ln\left(\frac{\Phi_i}{\theta_i}\right)\right] \tag{2}$$

Here, x_i is the mole fraction of the component i in the liquid phase. Φ_i and θ_i are molecular volumes and surface fractions, respectively, and expressed as:

$$\frac{\Phi_i}{x_i} = \frac{r_i}{\sum_j x_j r_j} \tag{3}$$

$$\frac{\theta_i}{x_i} = \frac{q_i}{\sum_j x_j q_j} \tag{4}$$

Here r_i, and q_i, respectively, are molecular volume and surface that are calculated from the group volume (R_k) and area (Q_k) parameters:

$$r_i = \sum_k \upsilon_k^i R_k \tag{5}$$

$$q_i = \sum_k \upsilon_k^i Q_k \tag{6}$$

υ_k^i stands for the number of group k in molecule i. The values of R_k and Q_k are calculated from the Bondi method. The coordination number Z is set equal to 10 and is consistent with the coordination number of molecules in the liquid phase. The residual part of the activity coefficient is:

$$\ln \gamma_i^{res} = \sum_k \upsilon_k^i \left[\ln \Gamma_k - \ln \Gamma_k^i \right] \tag{7}$$

Γ_k is the activity coefficient of group k at the solution composition, and Γ_k^i is the activity coefficient of group k in the pure component i. In this model, the interaction parameter between the main groups (τ_{ij}) is expressed as:

$$\tau_{ij} = \exp\left(-\frac{a_{ij}}{T} \right) \tag{8}$$

where a_{ij} has been regressed using experimental VLE data. The UNIFAC-DMD[15] and NIST-UNIFAC[30] models have the same formulation, but with different experimental data to regress the interaction parameters. So, these two models have different values for interaction parameters. The combinatorial part of these models is given as:

$$\ln \gamma_i^C = 1 - \frac{\Phi_i'}{x_i} + \ln\left(\frac{\Phi_i}{x_i} \right) - \frac{Z}{2}\left[1 - \frac{\Phi_i}{\theta_i} + \ln\left(\frac{\Phi_i}{\theta_i} \right) \right] \tag{9}$$

$$\frac{\Phi_i'}{x_i} = \frac{r_i^{\frac{3}{4}}}{\sum_j x_j r_j^{\frac{3}{4}}} \tag{10}$$

Also, these models include a temperature dependency of the interaction parameters as follows:

$$\tau_{ij} = \exp\left(-\frac{a_{ij} + b_{ij}T + c_{ij}T^2}{T} \right) \tag{11}$$

The values of R_k and Q_k in these two models are obtained via regression of the experimental data.

The COSMO-RS method was developed by Klamt[33] and is based on the molecular orbital continuum solvation models. The COSMO-SAC model, then, was developed by Lin and Sandler based on the COSMO-RS model.[34] This model performs superior to the COSMO-RS model for almost all phase equilibria data types.[34] In consequence, this study focuses on COSMO-SAC as a pure predictive model, which will be introduced in the following. This model uses individual atoms as the building blocks for predicting phase equilibria instead of functional groups in the group contribution models. Only the chemical structure of the molecules is needed to apply the model. So, the COSMO-based models provide a considerably larger range of applicability than group-contribution methods. Although the formulation of the combinatorial term for the activity coefficient is almost as the original UNIFAC method, the values of r_i and q_i are calculated as:

$$r_i = \frac{V_i}{V_{eff}} \tag{12}$$

$$q_i = \frac{A_i}{A_{eff}} \tag{13}$$

Here V_i, and A_i are the molecular volume, and molecular surface of component i, respectively. V_{eff}, and A_{eff}, respectively, stand for the standard component volume and surface, and the values are 66.69 Å^3 and 79.53 Å^2. The residual term of the activity coefficient is calculated as:[34]

$$\ln \gamma_i = \frac{A_i}{a_{eff}} \sum_{\sigma_m} p_i(\sigma_m)\left[\ln \Gamma_S(\sigma_m) - \ln \Gamma_i(\sigma_m)\right] \tag{14}$$

with

$$\ln \Gamma_S(\sigma_m) = -\ln\left\{\sum_{\sigma_n} p_S(\sigma_n)\Gamma_S(\sigma_n)\exp\left[-\frac{\Delta W(\sigma_m,\sigma_n)}{kT}\right]\right\} \tag{15}$$

$$\ln \Gamma_i(\sigma_m) = -\ln\left\{\sum_{\sigma_n} p_i(\sigma_n)\Gamma_i(\sigma_n)\exp\left[-\frac{\Delta W(\sigma_m,\sigma_n)}{kT}\right]\right\} \tag{16}$$

where $\Gamma_S(\sigma_m)$ is the segment activity coefficient of segment σ_m in solvent mixture, $\Gamma_i(\sigma_m)$ is the segment activity coefficient of segment σ_m in component i. $p_S(\sigma_n)$ stands for the sigma

profile of the solvent mixture, and $p_i(\sigma_n)$ is the sigma profile of the component *i*. σ represents the surface charge density and $\Delta W(\sigma_m, \sigma_n)$ is the exchange energy between segments *m* and *n*. The percentage of average absolute relative deviation (%AARD) for pressure and temperature is used to compare the predictive power of the original UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models as:

$$\%\mathrm{AARD}(T) = \left(\frac{100}{\mathrm{NP}}\right)\sum_i \left|\frac{T^{\mathrm{Exp}(i)} - T^{\mathrm{Pred}(i)}}{T^{\mathrm{Exp}(i)}}\right| \tag{15}$$

$$\%\mathrm{AARD}(P) = \left(\frac{100}{\mathrm{NP}}\right)\sum_i \left|\frac{P^{\mathrm{Exp}(i)} - P^{\mathrm{Pred}(i)}}{P^{\mathrm{Exp}(i)}}\right| \tag{16}$$

Here NP stands for the number of data points, and Exp, and Pred are experimental and predictive temperature or pressure, respectively.

2.2. Used molecules

A total of 200 molecules were used in this study; the molecules were divided into different homologous including alkanes, cycloalkanes, alkenes, aromatics, aldehydes, esters, ethers, amines, nitriles, carboxylic acids, alkyl halides, carbon disulfide, mercaptans, etc, and bio-oil related molecules (water, alcohols, ketones, anisole, phenol, mequinol, etc). The name and CAS number of the 200 studied molecules are given in supporting information 1. Anisole, Mequinol, catechols, acetol, guaiacol, vanillin, syringol, syringaldehyde, acrolein, and linoleic acid are some of the examples of the used bio oil-related molecules in this research. The chemical structure of these molecules is shown in Figure 1.

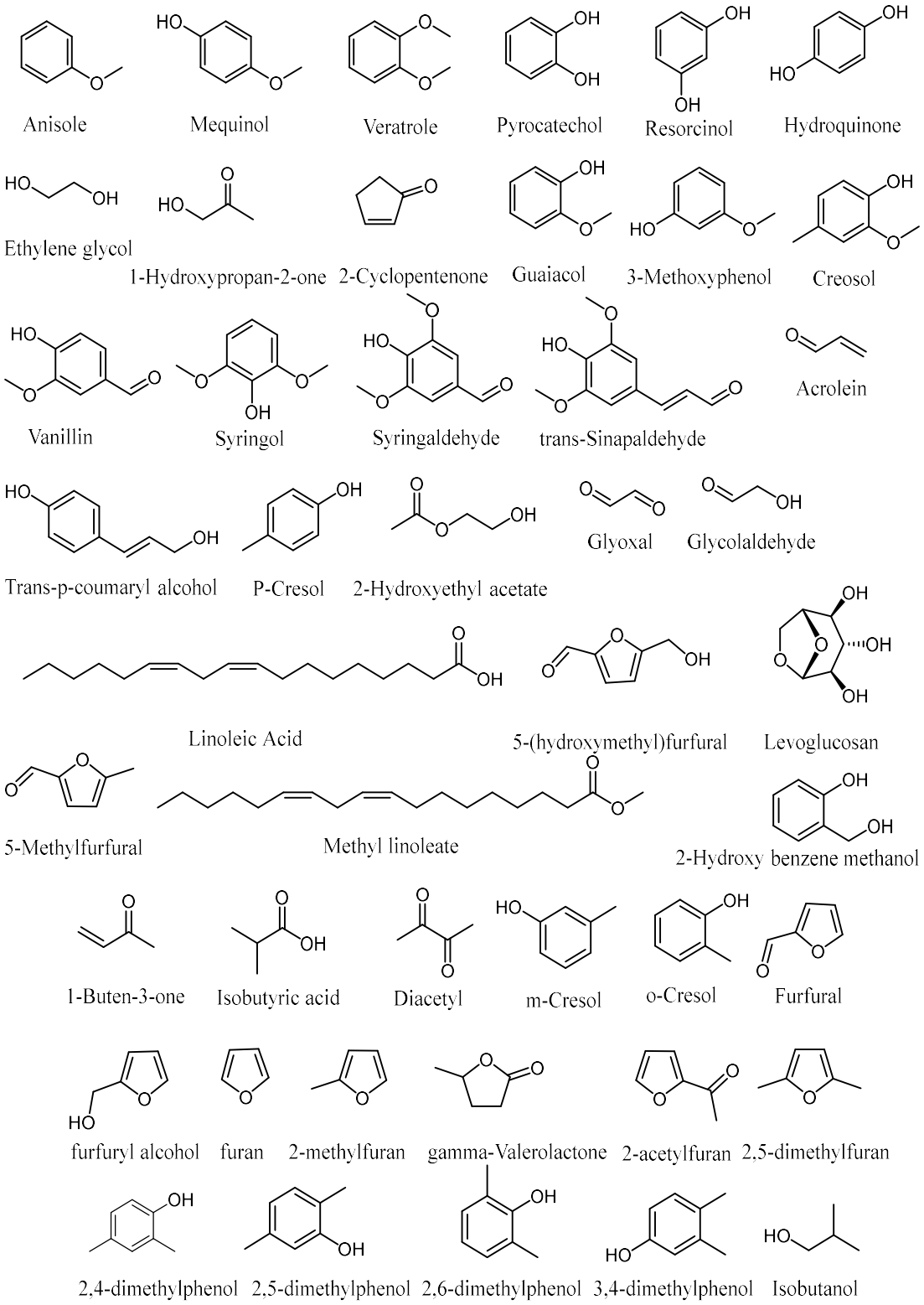


Figure 1: The chemical structure of the used bio-oil-related molecules

3. RESULTS and DISCUSSION

In this section, the results of the original UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC for the VLE of binary and ternary mixture, distillation curve, and LLE of binary and ternary mixture will be discussed.

3.1. Prediction of VLE for Binary Mixtures

The studied molecules for VLE prediction by the predictive models were classified into two classes including class I (alkanes, cycloalkanes, alkenes, aromatics, alcohols, ketones, amines, acids, chlorinated molecules, sulfur-containing molecules, nitrated molecules, halogenated) molecules, and class II (bio-oil related) molecules.[4] The predicted results by the models were compared with the experimental data. These data are categorized into isothermal data (variable pressure data) and isobar data (variable temperature data). Bubble pressure calculation is carried out for the former and bubble temperature calculation is done for the latter.

Table 1: The number of experimental isothermal or isobaric data sets used to study the predictive power of the UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models

Class		NO. of isothermal or isobar used data set
I: Including non-bio-oil molecules	Alkanes	54
	Cycloalkanes	38
	Alkenes	16
	Aromatics	270
	Amines	323
	Halogenated molecules	484
	Sulfur containing molecules	143
	Nitrated molecules	194
	Sundries molecules	92
	Total data set	2926 (47.88%)
II: Bio-oil molecules	Alcohols	668
	Ketones	551
	Carboxylic Acids	93
		phenol (32), anisole (43), mequinol (3), 1,2-dimethoxybenzene (7), catechol (20), ethylene glycol (75), guaiacol (16), acrolein (7), p-cresol (35), 2-hydroxyethyl acetate (3), isobutyric acid (7), diacetyl (16), m-cresol (30), o-cresol (51), furfuryl alcohol (11), furan (19), 2-methyl furan (20), γ-valerolactone (12), acetyl furan (4), 2,5-dimethylfuran (16), 2,4-xylenol (6), 2,6-xylenol (6), 3,4-xylenol (1), 3,5-xylenol (5)
	Total data set	445 (52.12%)

The deviation of four models based on the %AARD for pressure and temperature is provided in

Table 2. Similar to Table 1, the results are categorized into two main classes. For binary mixtures containing bio-oil and non-bio-oil molecules, the NIST-UNIFAC model provides the least deviation, while the COSMO-SAC gives the most deviation. Also, the difference between the deviation of the NIST-UNIFAC (%AARD in pressure=3.83, %AARD in temperature=0.30) and UNIFAC-DMD models (%AARD in pressure=3.78, %AARD in temperature=0.32) are low enough that it is not possible to distinguish between them (for non-bio-oil molecules). Based on the obtained results, in about half of the systems, the NIST-UNIFAC performs better than the UNIFAC-DMD model and vice versa. So, for these types of molecules, both the NIST-UNIFAC and UNIFAC-DMD are recommended to be used in the conceptual design of processes. These two models use a wide range of high-quality data in the interaction parameter optimization procedure, including VLE, LLE, SLE, h^{ex}, C_P^{ex}, etc. The original UNIFAC model has been fitted by Fredenslund et al.[23] only to the VLE data set. The results of the models for bio-oil-related molecules are very different. Based on the average deviation for both pressure and temperature, the original UNIFAC model with %AARD in pressure=11.08 and %AARD in temperature=0.83 performs superior to the other three models. This model provides the best results in seven cases. The NIST-UNIFAC model ranks second in terms of performance (%AARD in pressure=11.71 and %AARD in temperature=0.87). This model gives the lowest deviation in eight cases. The UNIFAC-DMD model with %AARD in pressure=12.71 and %AARD in temperature=0.90 and the lowest deviation in three cases ranks third. Although, based on the average deviation, the COSMO-SAC gives the highest deviation, it provides the least deviation in six cases including acrolein, diacetyl, o-cresol, furan, γ-valerolactone, and 2,6-xylenol. No logical conclusion can be made about which model is better, but these differences can be attributed to some reasons. First, the number of experimental datasets that have been used in the tuning interaction parameters of

the group contribution models is limited. So, the interaction parameters of the groups in class II are not as reliable as in class I. Second, the NIST-UNIFAC and UNIFAC-DMD have a larger number of parameters than the original UNIFAC model leading to an overfitting of the parameters with a potential negative effect on prediction results. In light of these, the COSMO-SAC model with not have much difference compared to other models, and because only the chemical structure is needed, is recommended to be used for the VLE of the binary mixture at least containing one bio-oil (class II) molecule in the mixture. The accuracy of the COSMO-SAC model for the predictions of the VLE may be improved by changing the molecular conformations that are used to generate the σ profiles.[35]

Table 2: Deviation of UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC based on the percentage of average absolute deviation (%AARD) for pressure and temperature for binary VLE systems

		UNIFAC		UNIFAC-DMD		NIST-UNIFAC		COSMO-SAC	
Class		%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)
I: Including non-bio-oil molecules	Alkanes	1.29	0.11	1.18	0.10	1.17	0.10	1.35	0.11
	Cycloalkanes	1.29	0.11	0.88	0.08	0.92	0.08	1.30	0.11
	Alkenes	0.94	0.08	0.87	0.07	0.98	0.08	1.11	0.09
	Aromatics	1.72	0.14	1.42	0.12	1.53	0.13	2.22	0.18
	Amines	5.41	0.53	3.52	0.40	3.63	0.32	7.30	0.70
	Halogenated molecules	3.84	0.33	2.86	0.25	2.80	0.25	7.44	0.67
	Sulfur containing molecules	6.37	0.51	5.95	0.51	5.18	0.42	13.12	1.12
	Nitrated molecules	11.25	0.91	10.93	0.78	8.72	0.81	14.47	1.39
	Sundries molecules	10.23	0.94	6.42	0.53	9.50	0.50	5.68	0.48
	Average	4.70	0.41	3.78	0.32	3.83	0.30	6.00	0.54
II: Bio-oil molecules	Alcohols	5.15	0.38	4.13	0.28	3.55	0.27	9.42	0.75
	Ketones	4.46	0.41	3.54	0.34	4.27	0.35	6.26	0.53
	Carboxylic Acids	9.34	0.67	7.70	0.59	6.02	0.44	11.2	0.92
	Phenol	4.06	0.34	4.16	0.35	4.82	0.47	8.20	0.79
	Anisole	6.76	0.53	5.12	0.43	4.88	0.40	5.16	0.41
	Mequinol	1.83	0.11	1.81	0.11	2.03	0.13	4.58	0.29
	1,2-Dimethoxybenzene	6.17	0.44	13.32	1.13	9.19	0.78	21.63	1.95
	Catechol	11.09	0.66	10.81	0.65	10.38	0.65	14.92	1.22
	Ethylene glycol	19.47	2.27	17.44	2.11	17.00	2.11	21.33	2.53
	Guaiacol	21.61	1.70	19.10	1.6	18.71	1.67	40.71	4.88
	Acrolein	9.01	0.72	19.03	1.58	20.36	1.73	7.96	0.68
	2-Hydroxyethyl acetate	11.31	0.67	26.85	1.47	17.95	1.02	32.46	4.01
	Isobutyric acid	5.27	0.37	5.18	0.37	4.45	0.31	11.28	0.78
	Diacetyl	27.49	1.46	34.52	1.81	33.12	1.75	3.44	0.28
	p-Cresol	16.75	1.65	14.83	1.50	14.56	1.64	15.47	1.49
	m-Cresol	5.87	0.50	4.78	0.40	5.26	0.46	7.21	0.59
	o-Cresol	13.94	0.90	13.13	0.82	11.33	0.86	9.95	0.73
	Furfuryl alcohol	11.68	0.64	15.98	0.81	15.67	0.79	12.95	0.99
	Furan	44.49	2.98	51.63	2.33	51.86	2.46	41.08	2.07
	2-Methyl furan	6.42	0.61	4.20	0.38	6.82	0.60	-	-
	γ-Valerolactone	12.72	0.92	20.28	1.33	15.15	1.13	8.96	0.77
	Acetyl furan	15.49	0.87	15.40	0.87	10.34	0.61	-	-
	2,5-Dimethylfuran	6.22	0.55	5.61	0.50	4.77	0.41	-	-
	2,4-Xylenol	11.37	1.39	12.47	1.70	11.49	1.53	15.83	2.20
	2,6-Xylenol	2.94	0.20	3.36	0.23	3.15	0.22	2.29	0.16
	3,4-Xylenol	0.61	0.05	0.95	0.08	1.17	0.10	1.03	0.08

	3,5-Xylenol	7.73	0.53	7.81	0.54	7.74	0.53	-	-
	Average	11.08	0.83	12.71	0.90	11.71	0.87	13.62	1.27

Figure 2 shows the %AARD in pressure of the four UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for all the used molecules in the binary VLE study. As is shown, the %AARD of all models is low enough for alkanes cycloalkanes, alkenes, and aromatics. This is due to the ideal/nearly ideal and moderate deviation systems from Raoult's law. The %AARD for the bio-oil-related molecules is given on the right-hand side of Figure 2, and as is shown deviation of the all four models almost for all studied molecules, particularly for furans, is higher than the non-bio-oil-related molecules. The reason for these results is given previously and here the results are shown graphically.

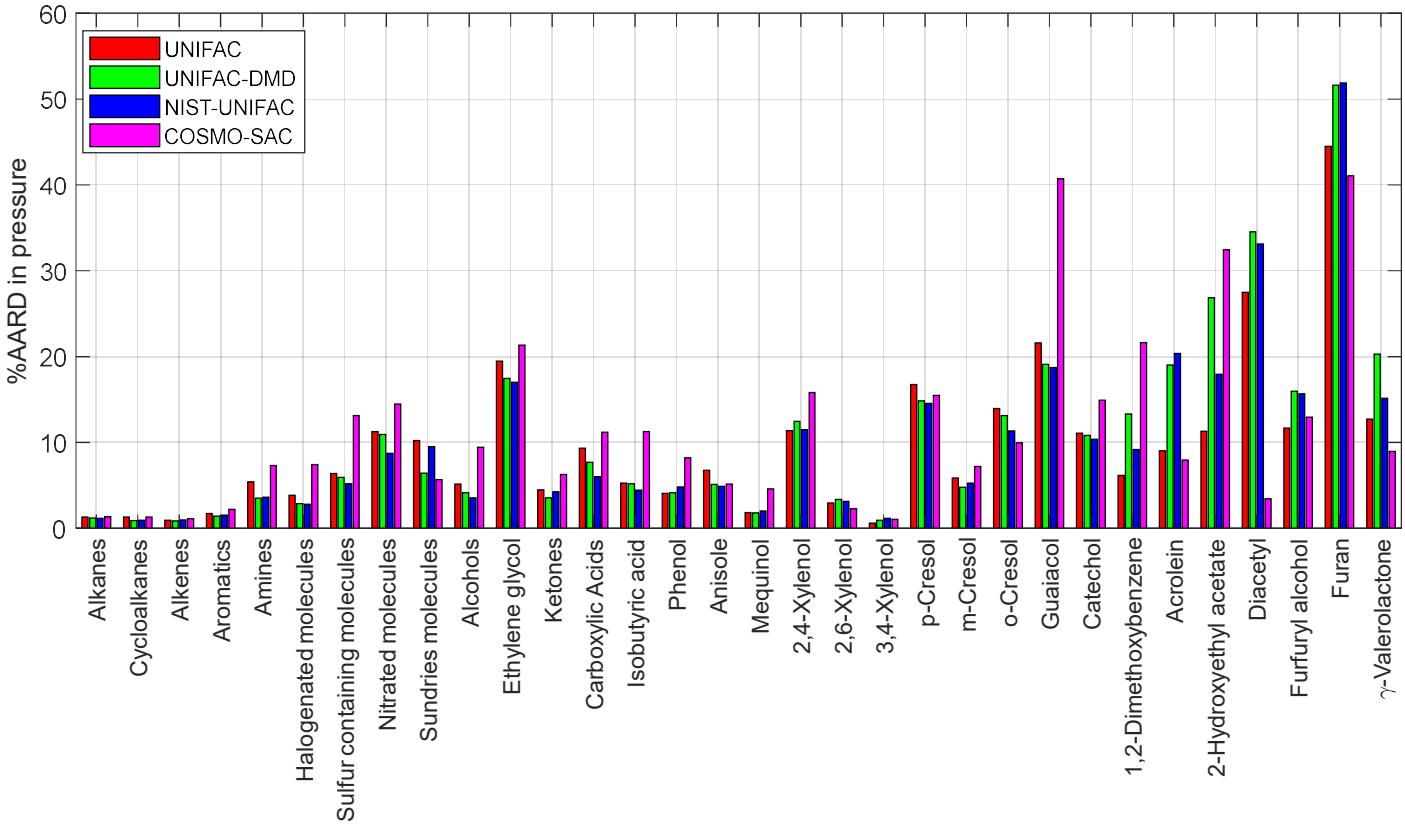


Figure 2: The %AARD in pressure of the four UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for all the used molecules in the binary VLE study

3.2. Prediction of VLE for Ternay Mixtures

Of the 200 molecules that are used in section 3.1 for the binary systems, to reduce the number of three molecule combinations ((200×199×198)/6) only 75 bio-oil related molecules (supporting information 2) were selected to be used in the ternary VLE calculation with 67525 ((75×74×73)/6) possible combinations. To the best of our knowledge, of 67525 combinations, only 74 ternary VLE systems with experimental data exist, but the quality of the 16 ternary systems was not enough. So, only 62 combinations (with 79 VLE datasets) remained to test the predictive power of the models for ternary VLE systems. Both bubble pressure and bubble temperature were used in the calculations. Table 3 provides the deviation of UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC based on the percentage of average absolute deviation for pressure (%AARD(P)) and temperature (%AARD(T)) for ternary VLE systems. In summary, %AARD(P)=7.82 and %AARD(T)=0.67 for UNIFAC, %AARD(P)=5.81 and %AARD(T)=0.48 for UNIFAC-DMD, %AARD(P)=5.89 and %AARD(T)=0.49 for NIST-UNIFAC, and %AARD(P)=10.48 and %AARD(T)=0.83 for COSMO-SAC were obtained. On the other hand, the UNIFAC-DMD and NIST-UNIFAC have been found to provide the best results. The COSMO-SAC as the pure predictive model gives the largest deviation (%AARD(P)=10.48 and %AARD(T)=0.83) compared to the other three group contribution models. Figure 3 shows the experimental[36] data versus the predicted result by the COSMO-SAC and NIST-UNIFAC models for n-hexane (1)+benzene(2)+isopropanol(3) ternary system at 101 kPa. As is shown the NIST-UNIFAC model shows a good agreement between the experimental data while there is a large deviation between the experimental data and the COSMO-SAC model. Although the COSMO-SAC model has the largest average deviation, it has the lowest deviation for 12 systems compared to the other three group contribution models. From the CPU time point of view, the

COSMO-SAC model requires the largest time to run these 62 ternary combinations in Table 3, so that it requires 51 minutes (CPU=Intel Core i9-12900k). In contrast, the UNIFAC, UNIFAC-DMD, and NIST-UNIFAC have an average CPU time equal to 14.7, 16.0, and 13.2 minutes. So, among these four models, the NIST-UNIFAC is the fastest model. It can be a challenging problem for multicomponent systems, in particular when the model is coupled with highly non-linear, thermodynamics, energy balance, mass balance, and other equations in process simulation software. It should be noted that most of the available ternary systems (Table 3) are conventional non-bio oil molecules so that the group contribution models provide the best results.

Table 3: Deviation of UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC based on the percentage of average absolute deviation (%AARD) for pressure and temperature for ternary VLE systems

		UNIFAC		UNIFAC-DMD		NIST-UNIFAC		COSMO-SAC	
Combination	NO. of data point	%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)	%AARD(P)	%AARD(T)
1-2-3	37	2.71	0.25	2.43	0.23	1.81	0.17	7.90	0.73
1-2-6	38	0.87	0.08	1.17	0.11	0.86	0.08	3.46	0.33
1-2-10	7	1.21	0.10	0.81	0.07	0.94	0.08	7.84	0.71
1-2-12	56	1.18	0.11	0.78	0.07	0.71	0.07	6.47	0.62
1-2-13	56	1.61	0.13	0.98	0.08	0.86	0.07	10.94	0.94
1-2-18	15	6.65	0.68	3.32	0.44	3.06	0.38	2.29	0.27
1-10-15	45	3.04	0.24	2.59	0.22	2.42	0.20	5.84	0.48
1-12-15	42	6.81	0.59	6.91	0.66	7.43	0.70	7.99	0.69
1-15-17	47	11.66	1.05	15.91	1.28	6.18	0.55	15.15	1.38
1-15-20	16	2.44	0.24	5.23	0.54	4.26	0.47	28.64	2.37
2-3-4	10	6.43	0.63	5.31	0.53	5.27	0.52	2.74	0.26
2-3-5	16	11.32	1.03	10.73	0.97	10.72	0.97	9.51	0.82
2-3-6	46	1.79	0.17	1.06	0.10	1.04	0.10	1.10	0.10
2-3-7	6	3.54	0.34	3.21	0.31	3.15	0.30	1.88	0.17
2-3-12	28	6.68	0.59	8.48	0.77	8.76	0.80	15.89	1.47
2-6-7	16	9.17	0.83	10.10	0.91	10.22	0.92	13.20	1.16
2-9-16	67	2.32	0.19	0.52	0.04	2.43	0.20	12.15	1.08
2-10-15	60	6.50	0.48	2.67	0.21	2.74	0.21	8.82	0.70
2-13-15	12	8.19	0.59	2.08	0.16	1.92	0.15	5.48	0.42
2-20-21	37	3.02	0.27	2.97	0.27	6.85	0.58	4.27	0.37
3-10-15	27	8.84	0.61	4.88	0.36	5.20	0.38	4.31	0.32
3-10-17	32	6.59	0.52	5.40	0.43	3.79	0.30	10.54	0.90
3-12-15	43	9.22	0.64	3.00	0.24	2.90	0.22	7.40	0.52
3-13-15	34	10.34	0.70	4.65	0.34	4.56	0.33	4.57	0.32
3-15-68	45	45.41	2.45	30.75	2.14	30.73	2.13	30.97	1.86
6-15-20	37	15.89	1.14	19.91	2.12	22.53	2.37	13.96	1.24
7-8-23	5	18.78	5.93	2.21	0.33	7.03	0.84	30.35	3.97
9-10-11	45	4.42	0.32	3.24	0.24	3.33	0.25	3.33	0.24
9-10-13	28	7.56	0.54	3.19	0.23	2.92	0.21	4.96	0.36
9-10-15	199	19.7	1.47	5.94	0.44	5.13	0.38	10.57	0.77
9-10-16	91	4.76	0.36	1.65	0.13	0.35	0.03	4.79	0.39
9-11-13	12	0.71	0.05	3.68	0.28	3.79	0.29	1.08	0.08
9-11-15	54	5.79	0.44	3.12	0.23	2.43	0.18	5.72	0.40
9-11-16	15	5.28	0.41	3.06	0.25	0.99	0.08	4.33	0.36
9-12-15	11	5.63	0.43	4.20	0.3	3.37	0.24	6.09	0.43
9-13-15	62	6.48	0.48	11.9	0.85	10.13	0.72	19.17	1.32
9-13-16	19	8.59	0.67	4.19	0.34	1.05	0.09	12.38	1.03
9-15-16	295	20.46	1.59	26.03	2.00	23.3	1.78	24.68	1.97
9-15-20	28	8.66	0.62	7.03	0.52	8.03	0.59	12.02	0.85
9-15-74	93	4.28	0.34	6.54	0.51	6.27	0.47	16.10	1.10
10-11-12	14	3.43	0.25	3.18	0.23	3.33	0.24	4.78	0.34
10-11-15	96	6.43	0.48	6.69	0.50	6.73	0.50	7.8	0.56
10-12-15	19	1.28	0.09	1.16	0.08	1.30	0.09	3.27	0.23
10-13-15	125	3.16	0.23	3.15	0.23	3.19	0.23	7.04	0.48
10-14-15	31	6.89	0.52	6.00	0.49	7.76	0.65	9.77	0.85
10-15-24	25	4.44	0.33	2.66	0.20	2.60	0.19	3.51	0.26
10-15-31	61	35.68	2.85	12.52	0.98	17.60	1.37	11.31	0.86
10-15-68	37	4.13	0.30	5.34	0.39	5.32	0.38	8.30	0.57
10-15-74	99	4.16	0.31	3.81	0.32	4.20	0.33	15.37	1.01
11-12-15	42	1.15	0.08	1.43	0.10	1.94	0.14	8.93	0.60
11-13-15	41	1.84	0.13	1.98	0.14	1.98	0.14	4.23	0.29
11-15-31	16	11.14	0.89	3.85	0.31	4.96	0.39	4.74	0.37
11-15-68	32	2.41	0.17	0.91	0.07	0.82	0.06	12.29	0.80
12-15-68	12	1.81	0.13	2.66	0.19	2.84	0.20	18.91	1.22
13-15-16	111	8.41	0.66	6.54	0.53	8.45	0.69	10.80	0.82
13-15-31	16	11.23	0.91	2.89	0.24	6.77	0.52	6.58	0.48
14-15-16	42	9.93	0.82	14.16	1.34	15.22	1.47	18.96	1.94
15-16-20	30	6.94	0.54	6.80	0.55	8.10	0.65	10.48	0.90
15-16-24	110	8.27	0.71	5.56	0.47	9.23	0.80	7.02	0.59
15-19-20	10	13.28	0.90	9.03	0.67	11.13	0.78	40.97	2.51
15-19-68	10	10.06	0.78	6.25	0.49	2.13	0.16	19.64	1.30
15-20-21	65	14.21	1.00	11.81	0.86	10.37	0.76	16.16	1.12

Total average: %AARD(P)=7.82 and %AARD(T)=0.67 for UNIFAC, %AARD(P)=5.81 and %AARD(T)=0.48 for UNIFAC-DMD, %AARD(P)=5.89 and %AARD(T)=0.49 for NIST-UNIFAC, and %AARD(P)=10.48 and %AARD(T)=0.83 for COSMO-SAC.

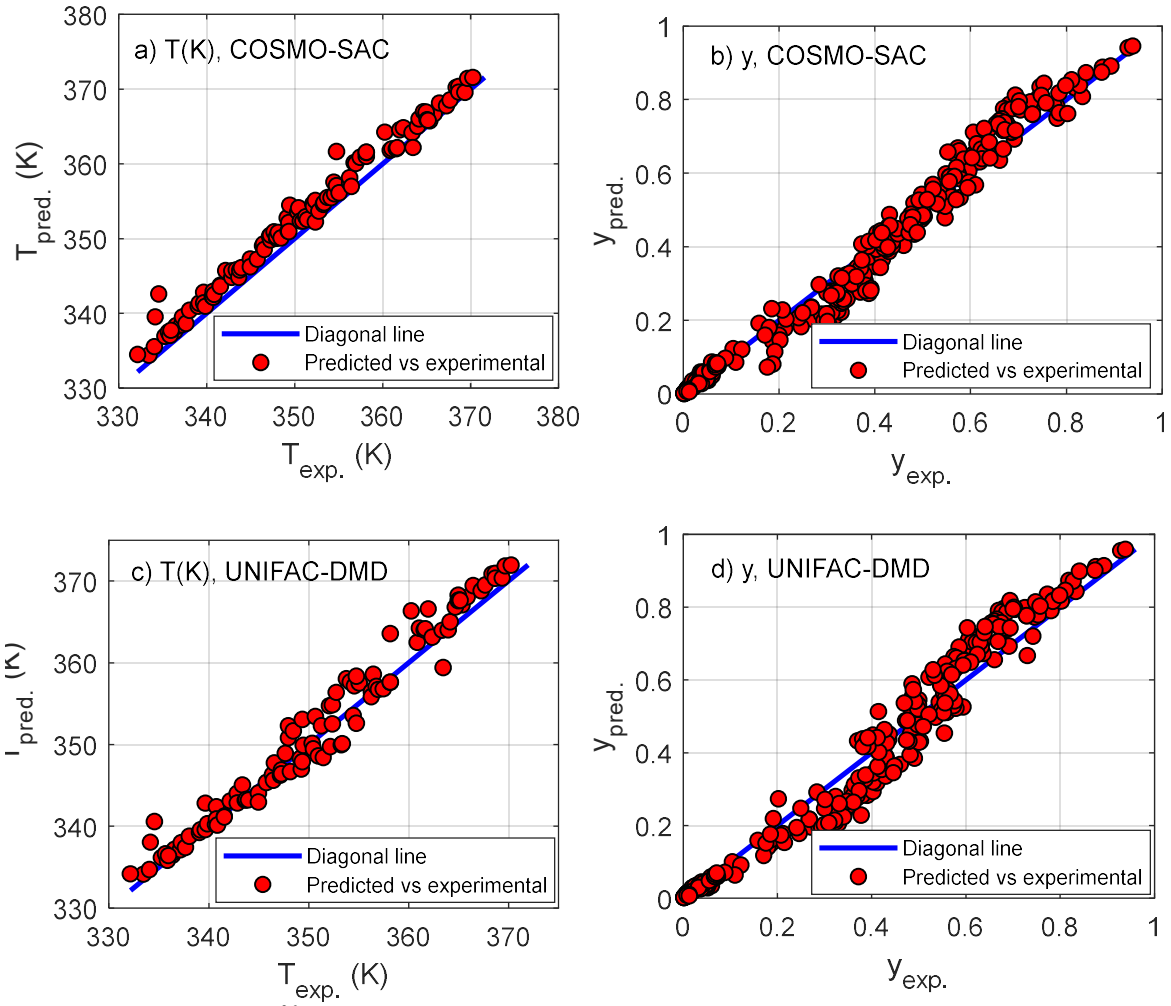


Figure 3: The experimental[36] data versus the predicted result by the COSMO-SAC and NIST-UNIFAC models for n-hexane (1)+benzene(2)+isopropanol(3) ternary system at 101 kPa

3.3.Prediction of LLE for Binary Mixtures

To find the binary LLE data 200 non-bio-oil and bio-oil related molecules were searched to find the experimental data. Finally, 125 binary systems (see supporting information 3) with 850 binary data sets and 2543 data points were gathered from the literature. Most of the systems form upper critical solution temperature (UCST) at the studied temperature range, while the systems of water+diisopropylamine, water+triethylamine, and water+n-propionaldehyde form lower critical solution temperature (LCST). Among the studied systems, the binary LLE system of water+methyl ethyl ketone (MEK) forms a phase envelope with UCST and LCST (Figure 4).

The required CPU time for 125 binary systems for UNIFAC. UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC are 17, 22, 22, and 6886 seconds (1.91 hrs). Although there is no significant deviation in CPU time among the three group contribution models, the UNIFAC model is the fastest with 17 seconds. The UNIFAC model cannot predict the phase split into two liquid phases for the binary systems of 5-116 (2-methyl-pentane+nitrobenzene[37], Figure 5a), 12-87, 50-51, 51-72, 51-87, 51-175, and 51-180 (red color cell in Table 4) while the experimental data show binary liquid-liquid systems. Also, the UNIFAC model cannot accurately predict the temperature dependency of the liquid phase mole fractions, in particular for systems where the liquid mole fraction changes considerably in terms of temperature (water+methyl ethyl ketone (MEK)[38-46], Figure 5b). Using the UNIFAC model, LCST was obtained for the systems 51-168 (water+p-cresol[47, 48], Figure 5c), 51-181, 51-182, 51-189, 51-190, 51-197, 51-192, and 51-191 while these models have UCST. The percent of average absolute deviation (%AAD) for the UNIFAC model for these 850 binary systems is 8.44.

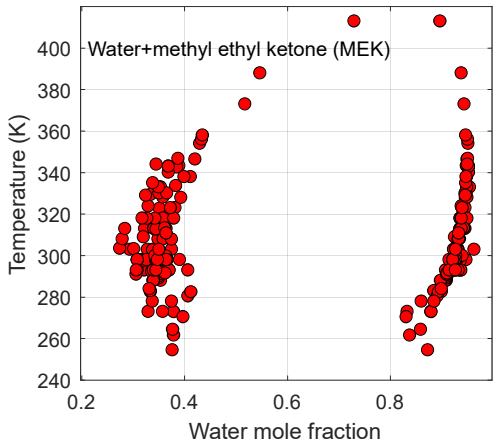


Figure 4: The experimental[38-46] binary LLE data for the system of water+methyl ethyl ketone (MEK)

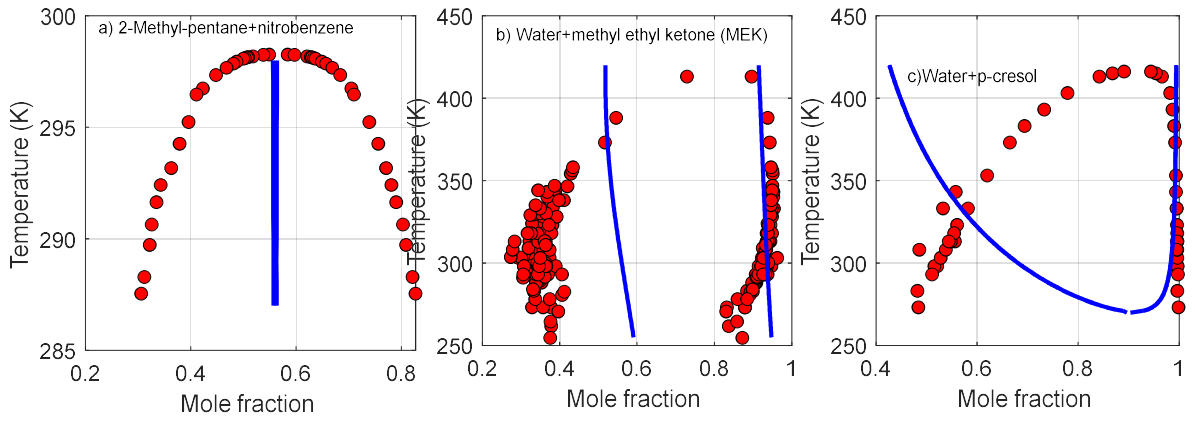


Figure 5: The experimental binary LLE data (●) and the predicted results by the UNIFAC model (▬) for the system of a) 2-methyl-pentane+nitrobenzene,[37] b) water+methyl ethyl ketone (MEK),[38-46] and c) water+p-cresol[47, 48]

The UNIFAC-DMD model shows a better performance than the UNIFAC model with a %AAD equal to 6.18. Also, this model cannot predict the phase split for the systems of n-hexane+ perfluorohexane (4-125), 2-methyl-pentane+nitrobenzene (5-116), n-heptane+perfluorohexane)[49] (9-125, Figure 6a), 2,2,4-trimethylpentane+ nitrobenzene (15-116), water+acrolein (51-166), and water+diacetyl (51-180). These systems are shown with green cells in Table 4. The UNIFAC-DMD model accurately predicts the phase behavior from the UCST and LCST points (water+methyl ethyl ketone[38-46], Figure 6b) and also the temperature dependency of the liquid phase mole fraction (water+m-cresol[47, 50], Figure 6c).

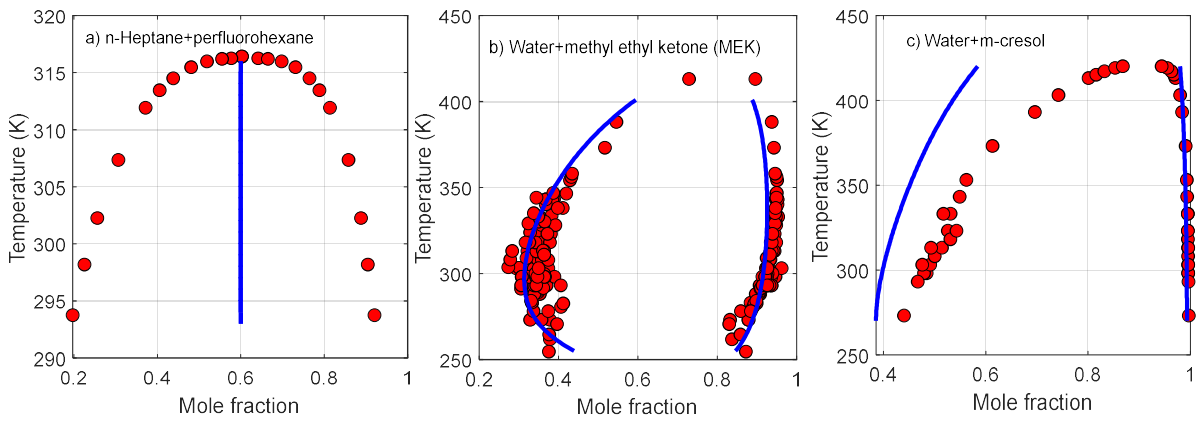


Figure 6: The experimental binary LLE data (●) and the predicted results by the UNIFAC-DMD model (▬) for the

system of a) n-heptane+perfluorohexane,[49] b) water+methyl ethyl ketone (MEK),[38-46] and c) water+m-cresol[47, 50]

Based on the percent of average absolute deviation (%AAD) given in Table 4, the NIST-UNIFAC model is the best model for the currently studied binary LLE systems, but similar to the UNIFAC, and UNIFAC-DMD models, the NIST-UNIFAC model could not predict the phase split (yellow color cell in Table 4) for 2-methyl-pentane+nitrobenzene (5-116, Figure 7a), n-octane+ dimethylamine (12-87), 2,2,4-trimethylpentane+nitrobenzene (15-116), toluene+perfluorohexane (29-125), water+diethylamine (51-79), water+dimethyl sulfide (51+111), and water+diacetyl (51+180) binary LLE systems. Of the temperature dependency of the liquid mole fractions, both can be predicted by the NIST-UNIFAC model (water+methyl ethyl ketone[38-46] (Figure 7a), and cyclohexane-methanol[41, 42] (Figure 7b)).

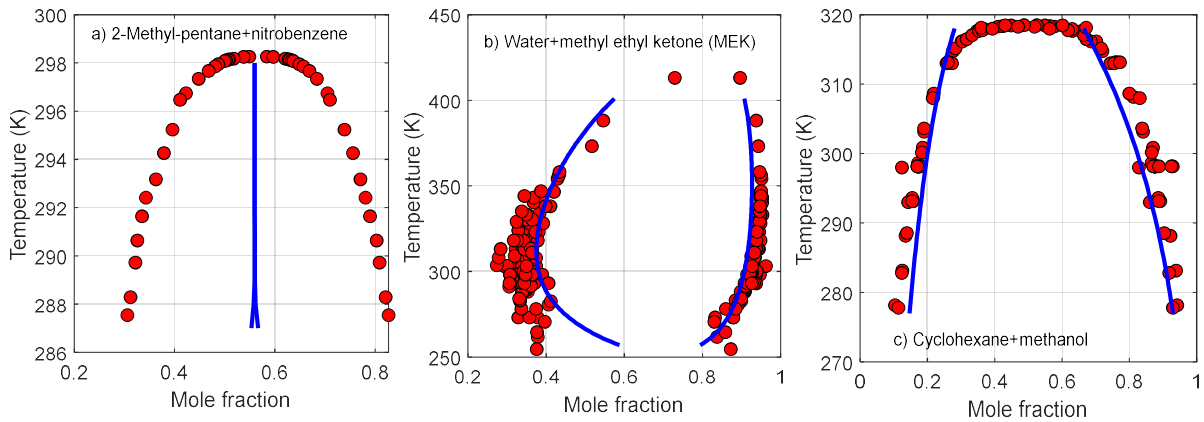


Figure 7: The experimental binary LLE data (●) and the predicted results by the NIST-UNIFAC model (—) for the system of a) 2-methyl-pentane+nitrobenzene,[37] b) water+methyl ethyl ketone (MEK),[38-46] and c) cyclohexane-methanol[51, 52]

The obtained results showed that the COSMO-SAC model with %AAD=12.46 cannot predict the phase split for a significant number of the studied systems. So, the results for this model are not provided in this section. Of particular interest is that none of the group contribution and pure predictive models can predict the phase split of the binary LLE systems of 2-methyl-pentane+2-nitropropane and water+diacetyl.

Table 4: Percent of average absolute deviation (%AAD) in liquid phase mole fraction for the UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for binary LLE systems

Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC	Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC	Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC
1-51	10	2.91	7.64	7.64	3.03	16-118	13	6.11	2.20	3.00	37.04	51-72	11	29.85	4.75	6.34	29.81
1-86	2	1.44	1.74	0.94	1.60	17-51	1	0.19	0.43	0.43	0.2	51-79	14	14.86	10.24	2.27	22.33
1-87	2	5.87	18.21	2.10	15.1	18-42	55	11.57	7.78	3.68	27.17	51-80	3	12.28	14.29	15.62	11.59
2-51	5	0.03	0.52	0.52	0.05	18-18	8	0.22	0.30	0.30	0.24	51-82	9	12.32	15.58	11.93	26.15
2-86	1	1.09	1.72	0.62	1.55	18-86	16	1.84	2.26	1.16	2.38	51-83	1	6.13	1.58	3.79	39.48
4-42	47	12.10	8.31	6.62	28.11	18-114	21	2.96	1.89	1.24	27.3	51-86	100	21.58	21.17	20.65	10.49
4-51	10	0.08	0.40	0.40	0.10	18-118	3	31.63	24.83	24.81	8.53	51-87	14	6.02	8.13	7.79	33.65
4-86	15	7.64	5.00	6.29	7.73	21-42	33	8.98	3.26	2.24	31.42	51-96	4	1.07	1.04	1.95	2.24
4-125	14	7.80	15.32	10.00	-	21-51	10	0.12	0.13	0.13	0.15	51-97	9	0.09	0.12	0.47	0.22
4-126	11	8.78	26.86	22.22	-	21-144	4	91.21	33.20	40.38	9.75	51-98	13	0.30	0.36	0.35	11.47
4-156	9	0.18	0.24	0.98	0.72	23-51	3	0.68	0.70	0.82	0.13	51-100	4	4.11	0.78	0.75	4.16
4-183	1	15.74	2.51	4.93	15.26	24-51	1	1.00	0.52	0.48	0.08	51-101	3	8.29	5.37	5.76	1.47
5-42	3	14.36	10.73	9.91	12.34	24-113	2	2.96	1.39	0.98	45.61	51-111	17	48.94	19.78	48.85	0.84
5-51	2	0.02	0.26	0.26	0.01	28-51	70	2.13	0.85	0.85	2.18	51-115	25	1.02	1.99	1.68	1.69
5-116	22	13.81	13.81	13.79	13.81	28-88	4	6.55	3.40	2.18	25.91	51-136	2	2.39	2.20	1.02	1.23
6-42	3	22.53	17.83	16.45	17.83	28-156	6	5.00	4.11	2.02	11.38	51-149	11	5.69	2.40	2.30	34.55
6-51	2	0.02	0.26	0.26	0.01	29-51	68	3.41	1.79	1.77	3.23	51-150	9	11.48	1.51	1.33	0.65
7-51	5	0.01	0.25	0.25	0.02	29-125	1	16.3	11.98	48.85	-	51-151	24	12.97	9.95	6.19	17.32
7-83	2	21.02	2.14	0.71	13.64	29-156	45	1.89	1.66	1.10	6.41	51-152	28	17.99	2.25	2.82	44.51
8-51	2	0.02	0.26	0.26	0.01	30-51	32	3.73	3.20	3.19	3.73	51-159	2	40.45	9.50	12.91	23.29
9-42	64	6.71	5.30	3.39	32.58	31-156	2	0.83	0.65	0.96	4.6	51-160	8	12.46	9.83	6.07	-
9-51	15	0.04	0.47	0.47	0.07	32-51	31	10.72	6.84	6.75	9.88	51-163	11	19.97	11.35	8.14	-
9-86	5	0.95	1.32	1.46	1.74	32-156	3	0.47	0.30	0.58	3.47	51-166	12	9.73	31.50	7.53	31.5
9-114	27	0.77	1.35	0.81	19.93	33-51	21	3.10	1.40	1.36	2.97	51-168	9	10.95	8.24	6.48	3.88
9-118	5	5.24	6.42	8.24	41.31	33-156	7	1.53	1.47	0.37	3.06	51-175	66	2.63	2.74	7.71	-
9-120	24	5.47	5.24	7.57	32.46	34-51	16	0.66	0.44	0.56	0.34	51-179	86	6.44	18.8	12.56	9.04
9-125	12	10.19	18.09	5.21	-	36-42	23	1.83	2.33	1.96	35.71	51-180	11	33.76	33.64	33.41	-
9-126	14	7.54	18.31	13.31	-	36-126	9	6.12	8.68	5.91	-	51-181	16	11.34	7.18	5.49	9.47
9-156	15	1.27	1.24	1.69	1.53	36-156	3	0.72	0.20	0.23	-	51-182	9	11.15	6.04	4.46	10.51
10-51	2	0.02	0.26	0.26	0.02	45-51	206	2.56	8.90	8.28	6.77	51-188	1	0.35	0.13	0.02	-
12-42	30	2.96	3.26	2.58	36.47	46-51	70	3.77	4.64	4.14	8.43	51-189	30	21.99	10.97	11.91	18.86
12-51	16	0.06	0.64	0.64	0.11	49-51	44	6.94	7.12	8.09	4.93	51-190	11	17.22	8.01	8.46	-
12-83	2	8.14	4.62	6.5	2.55	50-51	58	11.89	5.35	4.07	12.16	51-191	14	17.99	3.11	5.11	15.69
12-86	9	1.79	1.84	1.70	1.95	51-60	122	11.18	2.14	3.28	28.87	51-192	3	21.35	14.65	14.32	6.31
12-87	2	7.71	4.02	7.67	18.46	51-62	15	11.61	8.13	6.63	26.13	51-193	9	4.82	6.92	6.59	5.81
12-126	8	8.6	16.02	10.44	-	51-63	104	1.55	2.21	4.79	28.12	51-194	1	3.92	1.47	2.41	13.46
15-51	7	0.35	0.63	0.63	0.38	51-64	34	1.50	1.02	3.42	7.76	51-195	100	6.04	4.88	6.21	6.72
15-116	28	9.28	17.27	15.11	17.51	51-65	27	1.00	1.10	2.25	5.26	51-196	14	12.22	6.68	8.12	14.65
15-156	3	0.03	0.07	0.54	0.38	51-67	2	1.76	1.15	5.43	8.76	51-197	4	21.44	13.01	13.13	-
16-42	24	2.85	2.45	2.90	41.38	51-69	47	12.62	1.42	1.79	12.5	51-198	9	34.99	13.36	18.26	-
16-51	5	0.10	1.30	1.30	0.15	51-70	49	3.35	0.78	0.92	8.67	51-199	13	1.59	5.33	4.84	4.18
16-89	2	1.78	1.92	0.15	7.77	51-71	17	3.65	4.11	5.25	42.38						

$\%AAD = \left(\frac{100}{N}\right)\sum_i \left|x_i^{Exp.} - x_i^{Pred.}\right|$, %AAD for UNIFAC=8.44, %AAD for UNIFAC-DMD=6.18, %AAD for NIST-UNIFAC=6.08, and %AAD for COSMO-SAC=12.46.

3.4. Prediction of LLE for Ternary Mixtures

To the best of our knowledge, of 67525 combinations, only 172 ternary combinations (LLE systems) with experimental data were found, but the quality of the 14 ternary combinations was

not enough. So, 157 combinations containing bio-oil and non-bio-oil-related molecules were used to compare the activity coefficient models (supporting information 4). It should be noted that each combination may have several measured experimental ternary liquid-liquid phase equilibria data. These 157 combinations have 276 experimental ternary LLE data with 4145 data points. The percent of average absolute deviation (%AAD) of the liquid phase mole fraction for the UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for 81 ternary LLE systems is given in Table 5. The UNIFAC-DMD model with %AAD=7.22 provides the lowest deviation compared to the other three models. However, the difference between the results of the UNIFAC-DMD and NIST-UNIFAC (%AAD=7.50) is very low. Although based on the %AAD the UNIFAC model (%AAD=7.95) shows weak results compared to the UNIFAC-DMD and NIST-UNIFAC models, it gives better results for 70 systems than the UNIFAC-DMD model and 64 systems than the NIST-UNIFAC model. It must be emphasized that the missing interaction parameters in the group contribution models were replaced with zero. Similar to the binary and ternary VLE systems, and also the binary LLE systems the COSMO-SAC model gives the highest deviation (%AAD=9.58). So, this model is recommended when the quality of the interaction parameters of the group contribution models is low, the interaction parameters are missing or molecules have groups that have not been defined in the group contribution models..

Table 5: Percent of average absolute deviation (%AAD) in liquid phase mole fraction for the UNIFAC, UNIFAC-DMD, NIST-UNIFAC, and COSMO-SAC models for ternary LLE systems

		%AAD						%AAD						%AAD Involved bio-oil-related molecules			
Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC	Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC	Combination	NO. of data point	UNIFAC	UNIFAC-DMD	NIST-UNIFAC	COSMO-SAC
1-2-9	23	18.68	12.12	10.23	6.03	2-13-15	153	35.85	36.32	35.18	7.16	14-15-68	29	6.88	4.43	5.02	6.00
1-2-23	83	9.53	7.78	7.86	15.71	2-14-15	10	3.91	1.49	0.94	1.04	14-15-69	54	7.94	6.17	6.09	1.77
1-2-31	7	0.56	0.59	1.45	26.26	2-14-31	8	2.07	3.62	2.65	26.29	15-16-24	11	3.44	2.04	5.29	5.10
1-2-58	4	16.70	14.81	16.11	39.07	2-15-16	24	4.44	2.09	3.87	7.74	15-16-74	7	3.94	4.43	3.91	-
1-3-15	2	0.01	0.27	0.27	49.60	2-15-17	11	6.77	7.38	6.15	6.23	15-17-60	27	1.60	3.93	3.32	-
1-3-23	43	15.36	13.12	10.10	64.66	2-15-19	14	3.84	5.26	1.98	18.07	15-17-74	20	4.91	7.73	6.34	-
1-3-24	5	32.88	24.35	26.21	30.31	2-15-20	11	2.78	2.75	2.45	5.78	15-20-24	5	8.46	2.10	2.54	2.78
1-4-23	5	4.02	3.55	2.20	49.71	2-15-21	6	1.83	1.20	1.53	5.08	15-20-68	24	8.71	11.33	18.89	5.08
1-4-24	4	11.02	15.82	15.65	41.30	2-15-31	10	1.22	1.25	0.73	1.26	15-20-69	24	1.85	6.64	6.48	11.55
1-5-58	4	19.83	4.70	6.33	69.24	2-15-68	26	3.02	3.73	2.92	32.08	15-21-24	61	12.46	22.96	27.96	4.70
1-6-23	10	9.21	8.78	6.52	62.64	3-6-15	9	0.11	0.16	0.16	0.16	15-21-69	10	2.18	5.85	8.35	6.45
1-6-24	5	15.91	6.32	5.65	38.26	3-8-9	8	0.82	1.19	1.18	1.79	15-24-50	80	8.08	11.78	14.15	4.17
1-9-13	17	5.98	4.81	3.40	12.10	3-9-15	71	7.10	7.08	6.95	3.76	15-24-69	7	21.26	11.17	15.05	-
1-9-15	134	6.94	4.77	6.15	30.10	3-10-15	160	43.28	31.52	37.55	16.90	15-24-70	8	1.84	2.05	2.82	2.47
1-9-16	23	19.04	12.99	11.47	49.77	3-11-15	11	3.43	4.38	4.36	3.80	15-28-69	114	22.00	7.49	7.57	4.60
1-9-18	11	8.49	1.57	1.96	8.73	3-12-15	18	2.10	4.73	4.12	1.74	15-28-71	33	17.95	7.41	10.57	4.82
1-10-15	161	16.60	14.91	14.26	22.02	3-13-15	30	9.85	9.58	8.75	5.85	15-29-69	72	15.13	5.36	6.41	2.97
1-10-18	9	19.58	3.16	4.32	22.04	3-14-15	41	30.27	23.93	22.47	1.62	15-29-71	30	17.67	7.63	10.59	4.70
1-12-15	12	0.59	1.99	1.70	3.04	3-14-31	40	5.36	5.56	8.15	12.86	15-30-69	60	5.70	4.98	4.51	1.49
1-12-31	35	3.85	3.32	3.42	1.05	3-15-16	10	3.69	1.72	2.91	1.08	15-30-71	33	17.40	6.08	8.02	2.56
1-15-17	29	13.27	5.04	8.40	3.90	3-15-17	12	6.92	4.61	6.61	3.61	15-34-69	57	14.95	2.63	1.82	4.53
1-15-20	10	0.98	1.07	1.08	7.64	3-15-20	105	7.53	6.22	5.84	2.28	15-48-68	15	1.23	5.05	4.17	-
1-15-31	28	1.66	1.57	1.73	7.51	3-15-21	31	18.64	15.36	15.8	7.44	15-48-69	42	17.54	5.21	10.91	4.30
1-15-74	21	0.26	1.24	1.05	10.79	3-15-23	7	6.21	5.00	7.99	11.74	15-56-69	57	4.30	3.89	2.37	8.53
1-22-23	13	11.30	9.23	11.01	14.34	3-15-24	16	0.18	0.03	0.06	0.38	15-57-69	30	1.01	3.06	2.08	-
1-24-59	22	23.11	4.64	15.52	5.16	3-15-26	16	5.97	2.64	1.71	3.41	15-58-60	5	4.42	5.15	4.79	-
2-8-9	9	2.29	2.10	1.57	3.46	3-15-31	11	0.97	1.02	0.37	0.54	15-67-69	40	2.07	1.12	0.35	4.25
2-9-15	75	16.60	13.43	13.48	10.95	3-15-32	13	2.38	3.25	3.13	3.96	15-69-72	40	1.38	1.62	0.58	-

$\%\text{AAD} = \left(\frac{100}{N}\right)\sum_i \left|x_i^{\text{Exp.}} - x_i^{\text{Pred.}}\right|$, %AAD for UNIFAC=7.95, %AAD for UNIFAC-DMD=7.22, %AAD for NIST-UNIFAC=7.50, and %AAD for COSMO-SAC=9.58.

4. CONCLUSION

A very comprehensive study was conducted to evaluate the predictive power of predictive group contribution and quantum chemistry-based models. An extensive amount of experimental binary and ternary vapor-liquid and liquid-liquid equilibrium data containing bio-oil-related molecules and normal molecules were checked. Based on the obtained results for vapor-liquid equilibria systems, group contribution models are the best option for mixtures without bio-oil molecules, but for mixtures containing bio-oil species, these models should be used with caution, or the interaction parameters of these models to be re-fitted to the experimental data. The COSMO-SAC model can be a good alternative for the prediction of the VLE of mixtures containing bio-oil molecules. For liquid-liquid systems, all models have limitations in predicting the two-phase region (phase split), but this limitation is very serious for the COSMO-SAC model, especially in the case of binary systems, and this model is not recommended in any way. It also has a very long CPU time. Predictive models provide acceptable results for binary and ternary liquid-liquid systems. The UNIFAC-DMD and NIST-UNIFAC models give the best results from the CPU time and accuracy point of view and as a result, these models are recommended for the design of bio-oil involved units, at least in the conceptual design stage.

■ ASSOCIATED CONTENT

Supporting Information

This information is available free of charge via the Internet at ???.

Supporting Information S1:

Supporting Information S2:

Supporting Information S3:

Supporting Information S4:

■ AUTHOR INFORMATION

Authors

Amir Jalalinejad - *Institute of Catalysis Research and Technology (IKFT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany.* *orcid.org/0000-0002-8188-2054*

Jaber Yousefi Seyf - *Department of Chemical Engineering, Hamedan University of Technology, P.O. Box: 65155-579, Hamedan, Iran. orcid.org/0000-0001-5624-1127*

Axel Funke *- Institute of Catalysis Research and Technology (IKFT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany; orcid.org/0000- 0002-8425-1346; Email: axel.funke@kit.edu*

Nicolaus Dahmen - *Institute of Catalysis Research and Technology (IKFT), Hermann-von-Helmholtz-Platz 1, D-76344 Eggenstein-Leopoldshafen, Germany.*

Complete contact information is available at:

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript. Conceptualization: A. J., J. Y. S., Methodology: A. J., J. Y. S., Investigation: A. J., J. Y. S., Visualization: A. F., Project administration: A. F., Supervision: A. F., N. D., Writing-original draft: A. J., J. Y. S., Writing-review and editing: A. F., N. D.

Notes

The authors declare no competing financial interest.

For Table of Contents Use Only

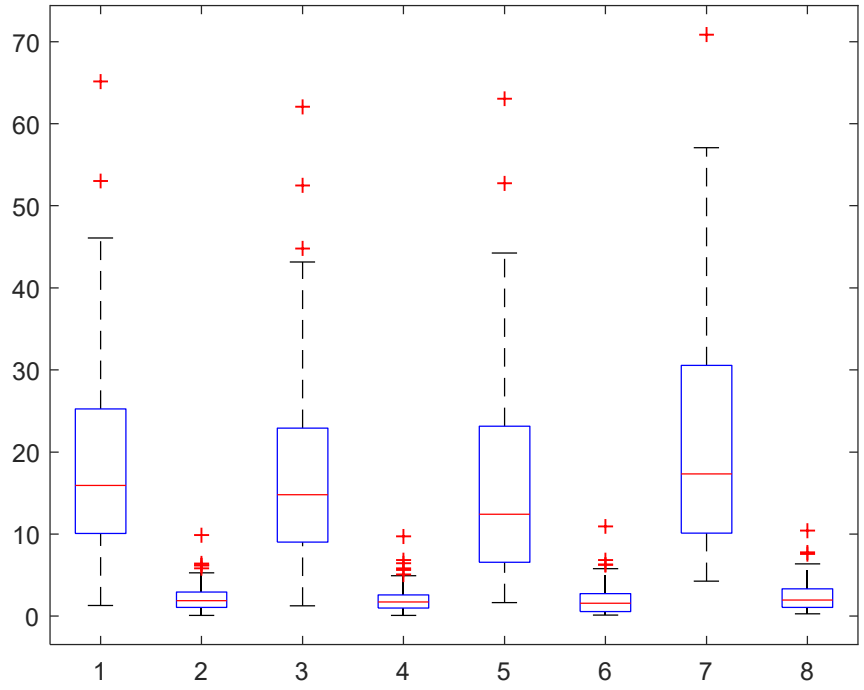
This paper deals with ???

(1) Cherubini, F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy conversion and management* **2010**, *51*, 1412-1421.
(2) Bridgwater, A. V. Renewable fuels and chemicals by thermal processing of biomass. *Chemical engineering journal* **2003**, *91*, 87-102.
(3) Czernik, S.; Bridgwater, A. Overview of applications of biomass fast pyrolysis oil. *Energy & fuels* **2004**, *18*, 590-598.
(4) Oasmaa, A.; Lehto, J.; Solantausta, Y.; Kallio, S. Historical review on VTT fast pyrolysis bio-oil production and upgrading. *Energy & Fuels* **2021**, *35*, 5683-5695.
(5) Oasmaa, A.; Sundqvist, T.; Kuoppala, E.; Garcia-Perez, M.; Solantausta, Y.; Lindfors, C.; Paasikallio, V. Controlling the phase stability of biomass fast pyrolysis bio-oils. *Energy & Fuels* **2015**, *29*, 4373-4381.

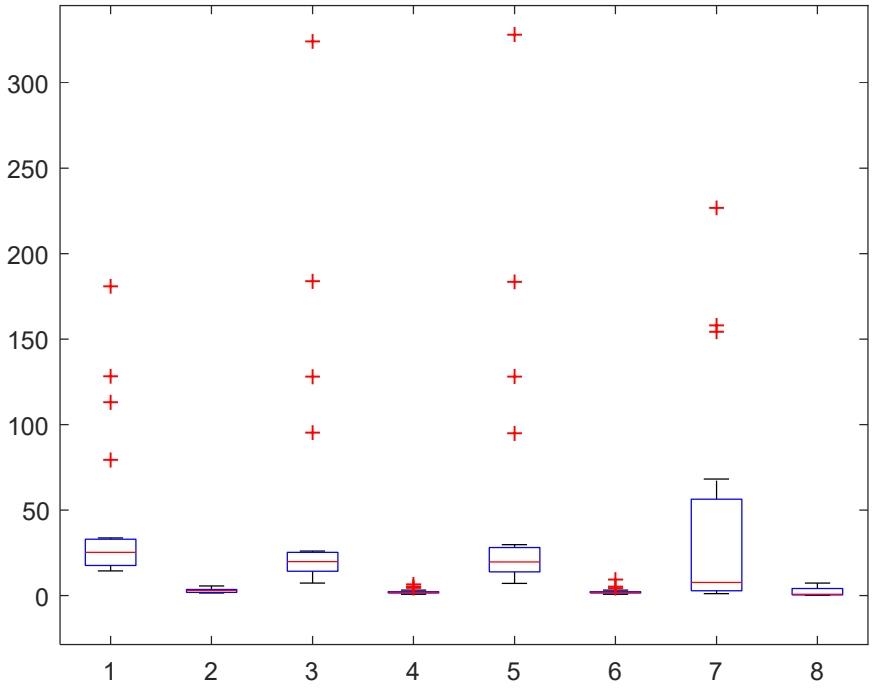
(6) Campos Fraga, M. M.; Vogt, J.; Lacerda de Oliveira Campos, B.; Schmitt, C. C.; Raffelt, K.; Dahmen, N. Investigation of Nb2O5 and Its Polymorphs as Catalyst Supports for Pyrolysis Oil Upgrading through Hydrodeoxygenation. *Energy & Fuels* **2023**, *37*, 10474-10492. DOI: 10.1021/acs.energyfuels.3c01152.
(7) Sun, Z.; Barta, K. Cleave and couple: toward fully sustainable catalytic conversion of lignocellulose to value added building blocks and fuels. *Chemical communications* **2018**, *54*, 7725-7745.
(8) Venderbosch, R. A critical view on catalytic pyrolysis of biomass. *ChemSusChem* **2015**, *8*, 1306-1316.
(9) Lindfors, C.; Elliott, D. C.; Prins, W.; Oasmaa, A.; Lehtonen, J. Co-processing of Biocrudes in Oil Refineries. ACS Publications: 2022.
(10) Pfitzer, C.; Dahmen, N.; Troger, N.; Weirich, F.; Sauer, J.; Gunther, A.; Muller-Hagedorn, M. Fast pyrolysis of wheat straw in the bioliq pilot plant. *Energy & Fuels* **2016**, *30*, 8047-8054.
(11) Niebel, A.; Funke, A.; Pfitzer, C.; Dahmen, N.; Weih, N.; Richter, D.; Zimmerlin, B. Fast pyrolysis of wheat straw—improvements of operational stability in 10 years of Bioliq pilot plant operation. *Energy & Fuels* **2021**, *35*, 11333-11345.
(12) Dahmen, N.; Abeln, J.; Eberhard, M.; Kolb, T.; Leibold, H.; Sauer, J.; Stapf, D.; Zimmerlin, B. The bioliq process for producing synthetic transportation fuels. *Wiley Interdisciplinary Reviews: Energy and Environment* **2017**, *6*, e236.
(13) Dahmen, N.; Sauer, J. Evaluation of techno-economic studies on the bioliq® process for synthetic fuels production from biomass. *Processes* **2021**, *9*, 684.
(14) Figueirêdo, M.; Hita, I.; Deuss, P.; Venderbosch, R.; Heeres, H. Pyrolytic lignin: a promising biorefinery feedstock for the production of fuels and valuable chemicals. *Green Chemistry* **2022**, *24*, 4680-4702.
(15) Griffin, M. B.; Iisa, K.; Wang, H.; Dutta, A.; Orton, K. A.; French, R. J.; Santosa, D. M.; Wilson, N.; Christensen, E.; Nash, C. Driving towards cost-competitive biofuels through catalytic fast pyrolysis by rethinking catalyst selection and reactor configuration. *Energy & Environmental Science* **2018**, *11*, 2904-2918.
(16) Iisa, K.; French, R. J.; Orton, K. A.; Dutta, A.; Schaidle, J. A. Production of low-oxygen bio-oil via ex situ catalytic fast pyrolysis and hydrotreating. *Fuel* **2017**, *207*, 413-422.
(17) De Hemptinne, J.-C.; Kontogeorgis, G. M.; Dohrn, R.; Economou, I. G.; Ten Kate, A.; Kuitunen, S.; Fele Žilnik, L.; De Angelis, M. G.; Vesovic, V. A view on the future of applied thermodynamics. *Industrial & Engineering Chemistry Research* **2022**, *61*, 14664-14680.
(18) Jalalinejad, A.; Seyf, J. Y.; Funke, A.; Dahmen, N. Solvent Screening for Separation of Lignin-Derived Molecules Using the NIST-UNIFAC Model. *ACS Sustainable Chemistry & Engineering* **2023**, *11*, 7863-7873. DOI: 10.1021/acssuschemeng.3c00906.
(19) Yousefi Seyf, J.; Asgari, M. Parametrization of PC-SAFT EoS for Solvents Reviewed for Use in Pharmaceutical Process Design: VLE, LLE, VLLE, and SLE Study. *Industrial & Engineering Chemistry Research* **2022**, *61*, 8252-8268. DOI: 10.1021/acs.iecr.2c00429.
(20) Wilson, G. M. Vapor-liquid equilibrium. XI. A new expression for the excess free energy of mixing. *Journal of the American Chemical Society* **1964**, *86*, 127-130.
(21) Renon, H.; Prausnitz, J. M. Local compositions in thermodynamic excess functions for liquid mixtures. *AIChE journal* **1968**, *14*, 135-144.
(22) Abrams, D. S.; Prausnitz, J. M. Statistical thermodynamics of liquid mixtures: a new expression for the excess Gibbs energy of partly or completely miscible systems. *AIChE journal* **1975**, *21*, 116-128.

(23) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE Journal* **1975**, *21*, 1086-1099.
(24) Weidlich, U.; Gmehling, J. A modified UNIFAC model. 1. Prediction of VLE, hE, and. gamma.. infin. *Industrial & engineering chemistry research* **1987**, *26*, 1372-1381.
(25) Gmehling, J.; Li, J.; Schiller, M. A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties. *Industrial & Engineering Chemistry Research* **1993**, *32*, 178-193.
(26) Gmehling, J.; Lohmann, J.; Jakob, A.; Li, J.; Joh, R. A modified UNIFAC (Dortmund) model. 3. Revision and extension. *Industrial & engineering chemistry research* **1998**, *37*, 4876-4882.
(27) Gmehling, J.; Wittig, R.; Lohmann, J.; Joh, R. A modified UNIFAC (Dortmund) model. 4. Revision and extension. *Industrial & engineering chemistry research* **2002**, *41*, 1678-1688.
(28) Jakob, A.; Grensemann, H.; Lohmann, J.; Gmehling, J. Further development of modified UNIFAC (Dortmund): Revision and extension 5. *Industrial & engineering chemistry research* **2006**, *45*, 7924-7933.
(29) Constantinescu, D.; Gmehling, J. r. Further development of modified UNIFAC (Dortmund): revision and extension 6. *Journal of Chemical & Engineering Data* **2016**, *61*, 2738-2748.
(30) Kang, J. W.; Diky, V.; Frenkel, M. New modified UNIFAC parameters using critically evaluated phase equilibrium data. *Fluid Phase Equilibria* **2015**, *388*, 128-141.
(31) Seyf, J. Y.; Flasafi, S. M.; Babaei, A. H. Development of the NRTL functional activity coefficient (NRTL-FAC) model using high quality and critically evaluated phase equilibria data. 1. *Fluid Phase Equilibria* **2021**, *541*, 113088.
(32) Seyf, J. Y.; Asgari, M.; Heidaryan, E. New interaction parameters from VLE data for group contribution (GC-NRTL) model. *Brazilian Journal of Chemical Engineering* **2022**, 1-12.
(33) Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *The Journal of Physical Chemistry* **1995**, *99*, 2224-2235.
(34) Lin, S.-T.; Sandler, S. I. A priori phase equilibrium prediction from a segment contribution solvation model. *Industrial & engineering chemistry research* **2002**, *41*, 899-913.
(35) Li, J.; Paricaud, P. Application of the Conduct-like Screening Models for Real Solvent and Segment Activity Coefficient for the Predictions of Partition Coefficients and Vapor–Liquid and Liquid–Liquid Equilibria of Bio-oil-Related Mixtures. *Energy & Fuels* **2012**, *26*, 3756-3768. DOI: 10.1021/ef300181j.
(36) Govindaswamy, S.; Andiappan, A.; Lakshmanan, S. ISOBARIC VAPOUR-LIQUID EQUILIBRIUM DATA FOR THE TERNARY AND SUB-BINARY SYSTEMS CONTAINING n-HEXANE (1)-BENZENE (2)-ISOPROPANOL-ISOPROPANOL (3). *Journal of Chemical Engineering of Japan* **1976**, *9*, 345-349.
(37) Du, Z.; Yin, T.; Xia, Y.; Cui, Z.; Shen, W. The liquid–liquid coexistence curves of {x nitrobenzene+(1− x) 2-methylpentane} and {x nitrobenzene+(1− x) isooctane} in the critical region. *Fluid Phase Equilibria* **2013**, *358*, 282-289.
(38) Ginnings, P.; Plonk, D.; Carter, E. Aqueous solubilities of some aliphatic ketones. *Journal of the American Chemical Society* **1940**, *62*, 1923-1924.
(39) Randall, M.; McKenna, F. The System Methyl Ethyl Ketone-Water below 0°. *Journal of the American Chemical Society* **1951**, *73*, 4859-4861.
(40) Campbell, A.; Kartzmark, E.; Falconer, W. The system: nicotine–methylethyl ketone–water. *Canadian Journal of Chemistry* **1958**, *36*, 1475-1486.

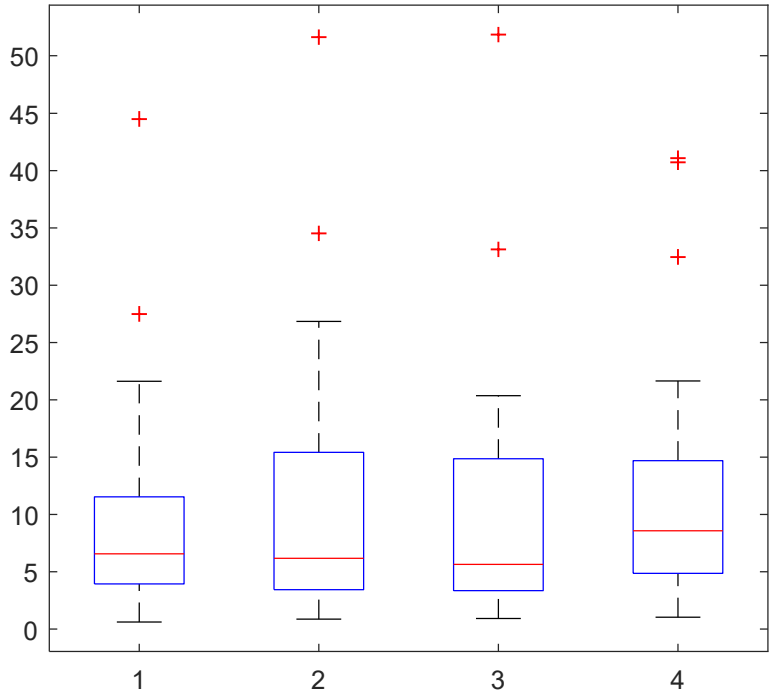
(41) Siegelman, I.; Sorum, C. Phase Equilibrium Relationships in the Binary System Methyl Ethyl Ketone–Water. *Canadian Journal of Chemistry* **1960**, *38*, 2015-2023.
(42) Altsybeeva, A.; Morachevskii, A. PHASE EQUILIBRIA AND THERMODYNAMIC PROPERTIES OF THE METHYL ETHYL KETONE-WATER SYSTEM. *ZHURNAL FIZICHESKOI KHIMII* **1964**, *38*, 1569-1573.
(43) Hunt, A.; Lamb, J. Liquid—liquid phase equilibrium in the 2-butanone—water system at high pressure. *Fluid Phase Equilibria* **1979**, *3*, 177-184.
(44) Ferino, I.; Marongiu, B.; Solinas, V.; Torrazza, S. Thermodynamic properties of aqueous non-electrolyte mixtures. II. Excess enthalpies and liquid—liquid equilibrium of 2-alkanone+ water mixtures. *Thermochimica acta* **1983**, *70*, 149-161.
(45) Ruiz, F.; Galan, M. I.; Prats, D.; Ancheta, A. M. Temperature influence on the ternary system 1-butanol-butanone-water. *Journal of Chemical and Engineering Data* **1984**, *29*, 143-146.
(46) Stephenson, R. M. Mutual solubilities: water-ketones, water-ethers, and water-gasoline-alcohols. *Journal of Chemical and Engineering Data* **1992**, *37*, 80-95.
(47) Erichsen, L.; Dobbert, E. The reciprocal relation of solubility of alkylphenols and water. *Brennst Chem* **1955**, *36*, 338-345.
(48) Li, H.; Wan, L.; Chu, G.; Tan, W.; Liu, B.; Qin, Y.; Feng, Y.; Sun, D.; Fang, Y. (Liquid+ liquid) extraction of phenols from aqueous solutions with cineole. *The Journal of Chemical Thermodynamics* **2017**, *107*, 95-103.
(49) Matsuda, H.; Hirota, Y.; Kurihara, K.; Tochigi, K.; Ochi, K. Liquid–liquid equilibria containing fluorous solvents as environmentally benign solvent. *Fluid Phase Equilibria* **2013**, *357*, 71-75.
(50) Leet, W. A.; Lin, H.-M.; Chao, K.-C. Mutual solubilities in six binary mixtures of water+ a heavy hydrocarbon or a derivative. *Journal of Chemical and Engineering Data* **1987**, *32*, 37-40.
(51) Katayama, H. Liquid–liquid equilibria of two ternary systems: methanol–cyclohexane including 1, 3-dioxolane or 1, 4-dioxane in the range of 277.79–308.64 K. *Fluid phase equilibria* **1999**, *164*, 83-95.
(52) Atik, Z.; Kerboub, W. Liquid− liquid equilibrium of (cyclohexane+ 2, 2, 2-trifluoroethanol) and (cyclohexane+ methanol) from (278.15 to 318.15) K. *Journal of Chemical & Engineering Data* **2008**, *53*, 1669-1671.



Box plot for ethylene glycol



Box plot for furanes



Box plot for furanes