Evolutionary optimization strategies for Liquid-liquid interaction parameters

Alessandro Robbiano[(a)], Antonio Tripodi[(b,*)], Francesco Conte[(b)], Gianguido Ramis[(a)], Ilenia Rossetti[(b,*)]

(a) DICCA, Università degli Studi di Genova, via all'Opera Pia 15A, 16100 Genoa, Italy;

(b) Chemical Plants and Industrial Chemistry Group, Dip. Chimica, Università degli Studi di Milano, via C. Golgi 19, 20133 Milano, Italy;

(*) corresponding authors: antonio.tripodi@unimi.it; ilenia.rossetti@unimi.it.

Abstract

A multi-level program for regressing thermodynamic parameters was tested for the UNIQUAC and NRTL models, on two type-I ternary mixtures showing liquid-liquid equilibria: n-heptane + benzene + methanol and water + ethanol + toluene. The presented algorithm is based on the isoactivity condition of coexisting phases, comprises sequential tests on the Gibb's energy shapes and uses an evolutionary strategy to obtain the parameters. Evaluating the outcome of 12 regression cases, it was found that: a) the unmodified UNIQUAC equation slightly outperforms the NRTL model for the given systems, b) the consistency check of the excess energy function acts as a bottleneck, and c) the use of different phases optimizer has a great impact on the overall performance.

Keywords:

Introduction

Thermodynamic models have long been a fundamental tool in chemical engineering for the prediction and automatic calculation of phase equilibria and heat transfer. Among them, the Non-Random Two Liquid [1] (NRTL) and Universal Quasi-Chemical [2] (UNIQUAC) expression for liquid-phase activity coefficients have gained importance and are widely used [3,4]. Since their first formulation, anyway, two problems appeared: first, finding out a mixture's composition at fixed parameters is a problem that cannot be solved directly, but requires recursive calculations [5]; second, the model parameters themselves must be calculated by a fitting procedure on the basis of experimental data (or other known parameters) [5].

Recently Marcilla et al. critically reviewed the regression of thermodynamic parameters from binary and ternary LLE data [6]. They pointed out that errors can arise both from limitations of the thermodynamic model and of the regression algorithm, so that an optimization should try a wide range of parameters to reduce the risk of being stuck in a local minimum [7]. Not any calculation procedure grants automatically the thermodynamic consistency of the parameters: optimization results based on the isoactivity condition (necessary but not sufficient), need to be tested for the stability of the examined equilibrium compositions [8]. For a ternary mixture involved in a Liquid-Liquid Equilibrium (LLE) problem, this criterion must be applied

also to the three binary mixtures virtually present. In fact, the way itself in which the activity coefficients are obtained (with the UNIQUAC and NRTL models) does not grant that calculated parameters yield a qualitatively correct ΔG_{mix} (respect to a chosen species molar fraction) for the binary systems, even if a tie-line between two ternary phases (along a direction in the phase diagram) does [9].

The interplay between the non-direct solution of a model's equations (which can be called the "inner shell" of such calculations) at fixed parameters and the evaluation and adjustment of the parameters themselves (the process "outer shell") has been first reviewed by Sorensen et al. [8] and, more recently, by Mitsos et al. [10]. The latter brought the multi-level recursive calculation and the ΔG_{mix} consistency under the same, up-to-date framework (for binary mixtures). While sophisticated mathematical procedures have been developed to solve simultaneously the phase-equilibrium and dG minimization [11,12], also simplified approaches proved fruitful [13].

While the problem of describing a ternary mixture on the basis of binary data and parameters, and of their consistency in general, has often been treated increasing the model's adjustable parameters or even modifying its underlying equations [1,9,14], later works focus rather on the capability of the algorithms to reach best-fitting *and* consistent solutions [15]. In a recent work, Tomassetti et al. [16] showed that the use of a genetic procedure with phase-stability tests yielded reliable results for type-I mixtures that present known criticalities. Their approach was based on a sequence of nested consistency checks, local optimizations in the phases' compositions space, and global optimization in the parameters space using an evolutionary procedure.

Optimization methods classified as "evolutionary strategies" (or "genetic") can work both as local and global optima seeker. With respect to other optimization methods, they have the following advantages:

a) their objective function must not be continuous, and there's no need to calculate its local Jacobian;
b) they do not need to be trained over extended datasets previously collected;
c) they are generally faster than Markovian-chain methods;
d) they are less easily trapped into local minima with respect to simplex-based (or similar) procedures, that often require a careful redefinition of the objective function to overcome this issue.

On the other hand, they are sensitive to the initial guess of the parameters and require several preliminary runs on already-known problems to fix the best combination of their many tunable features. For a detailed description of these kind of algorithms, the reader is referred to [17] and [18].

The scope of this work is to test and extend further the applicability of the algorithm recently proposed by Tomassetti et al. (with minor modifications), conducting a case study that involves both UNIQUAC and NRTL methods, two ternary mixtures and different approaches to deal with selected calculation aspects. Though the chosen mixtures (heptane+benzene+methanol and water+ethanol+toluene) belong to the same type-I of LLE from a thermodynamic point of view, their critical regions cover appreciably different regions: in particular, the water-toluene pair shows a miscibility gap practically as wide as the whole composition range, so the

numerical discretization of the composition space becomes an additional problem when seeking ΔG_{mix} minima in this case.

Methods

The optimization strategy adopted foresees the interplay of two shells [16]: the inner (right side of Figure 1) algorithm finds the coexisting phases that, at fixed interaction parameters, better approximate the experimental data; the outer algorithm (left section of Figure 1) seeks the parameters that give the better results in the inner sections. The capability of a given set of parameters to describe the system is then evaluated by three objective functions, based on the liquid-phase activity coefficient γ:

1. the activity of the three components $a_i = \gamma_i x_i$ in the two coexisting phases I, II at fixed fractions and variable parameters $\underline{\underline{U}}$:
   $OF_1\left(\underline{\underline{U}}\right) = \sum_{i=1}^{3}\sum_{d=1}^{data}(a_{i,d}^{I} - a_{i,d}^{II})^2$;
2. the proximity of the calculated phases compositions x to the reference ones $\bar{x}$ at fixed parameters:
   $OF_2(x) = \sum_{i=1}^{3}\sum_{d=1}^{data}\sum_{p=1}^{2}(x_{i,d}^{p} - \bar{x}_{i,d}^{p})^2$;
3. on the summation of these contribution:
   $OF_3 = (OF_1 + OF_2)^{1/2}$

Since the equality of the chemical potentials (i.e. of the activities) is a necessary but not sufficient condition, the parameters' consistency is checked by the binary mixtures ΔG_{mix} shape and by the Common Tangent Plane (CTP) test [19]. The main steps are then:

1. beginning of the outer shell algorithm (genetic algorithm), loading of the experimental coexistence data and a first guess for the thermodynamic parameters – a list of the evolutionary options most often used is given in the supporting info.
2. Optional parameters adjustment, calculation of OF_1.
3. Calculation of the Gibbs' free energy of mixing: $\Delta G_{mix} = \Delta G_e + \Delta^0 G$ for each binary mixture.
4. Check on the ΔG_{mix} shape with respect to the known miscibility of the components (if the check is not passed, return to the outer shell algorithm and initialization of a new individual).
5. Initialization of the first guess for two coexisting phases $\boldsymbol{x}^I$, $\boldsymbol{x}^{II}$.
6. Optimization of the isoactivity condition adjusting the calculated equilibrium compositions.
7. Common Tangent Plane (CTP) test on the calculated tie line: first the Tangent Plane Distance Function [19] is calculated for a discrete set of $\boldsymbol{x}_k$ compositions along the distance $|\boldsymbol{x}^I - \boldsymbol{x}^{II}|$:
   $$TPDF(\boldsymbol{x}_k) = \Delta G_{mix}(\boldsymbol{x}_k) - \left[\Delta G_{mix}(\boldsymbol{x}^{II}) + \frac{\Delta G_{mix}(\boldsymbol{x}^I) - \Delta G_{mix}(\boldsymbol{x}^{II})}{\boldsymbol{x}^I - \boldsymbol{x}^{II}}(\boldsymbol{x}_k - \boldsymbol{x}^{II})\right]$$
   then the values and positions of the $TPDF$ mimima are checked testing the inequalities:
   $TPDF(\boldsymbol{x}_k) \geq TPDF(\boldsymbol{x}^{I,II})\ \forall k$, and: $err \geq |TPDF(\boldsymbol{x}^{I,II})|$ where *err* is a tunable positive threshold.

8. If the test is passed, calculate OF_2 and move on to the next tie line, otherwise reinitialize the first guess of the phases and repeat the calculation of the tie line.
9. If the maximum number of iteration is reached move on to the next tie line and add a penalty on the final error.
10. Calculate OF_3 and check if the outer shell algorithm is converged.
11. If yes, stop the algorithm and save the final results; otherwise, start the next generation.

To obtain a tie-line, the maximum number of trials (steps 8 – 9) is varied between 200 and 400, while the threshold on the CTP condition (step 7) is varied between 10^{-4} and 10^{-2}. It must be pointed out that this is the algorithm section that requires the longer calculation time, so the execution of the test turns it into an actual computational bottleneck. Nevertheless, this slow-down of the "inner shell" is compensated from the fact that the "outer" genetic procedure does not waste generations dealing with non-consisted parameters. The tested cases are listed in Table 1, while the main algorithm parameters in Table 2. The systems examined are described in the works of García-Flores et al. [20] (mixture 1) and Gramajo de Doz et al. [21] (mixture 2).

Case	Thermodynamic model	Tie line calculation method	Mixture	Inner steps
1	UNIQUAC	Tommassetti	Heptane (1) + Benzene (2) + Methanol (3)	3
2	UNIQUAC	Sorensen	Heptane (1) + Benzene (2) + Methanol (3)	3
3	UNIQUAC	Sorensen	Heptane (1) + Benzene (2) + Methanol (3)	2
4	NRTL	Tommassetti	Heptane (1) + Benzene (2) + Methanol (3)	3
5	NRTL	Sorensen	Heptane (1) + Benzene (2) + Methanol (3)	3
6	NRTL	Sorensen	Heptane (1) + Benzene (2) + Methanol (3)	2
7	UNIQUAC	Tommassetti	Water (1) + Ethanol (2) + Toluene (3)	3
8	UNIQUAC	Sorensen	Water (1) + Ethanol (2) + Toluene (3)	3
9	UNIQUAC	Sorensen	Water (1) + Ethanol (2) + Toluene (3)	2
10	NRTL	Tommassetti	Water (1) + Ethanol (2) + Toluene (3)	3
11	NRTL	Sorensen	Water (1) + Ethanol (2) + Toluene (3)	3
12	NRTL	Sorensen	Water (1) + Ethanol (2) + Toluene (3)	2

Table 1: data, thermodynamic models, and tie lines search methods tested. Mixture 1 (Heptane-Benzene-Methanol) is referenced in the work of Hernandez & al., Mixture 2 (Water-Ethanol-Toluene) is taken from Gramajo & al.

The phases compositions actually managed during the calculus are not the experimental ones, but slightly modified copies [8,16], and are part of the optimized parameters, i.e. of an individual "genome". This allows the heuristic modification of possibly biased experimental data to essentially correct thermodynamic parameters, and is a useful feature in regions of the phase diagram where the calculated ΔG_{mix} minima are shallow or the activities are more sensitive to slight molar fractions' variations.

Step	Maximum iterations	Adjusted variables	Algorithm	Tolerance	Objective function	Test applied on the results

Isoactivity optimization	1000	Method parameters	Nelder-Mead	10^{-5}	OF_1	Check of the Gibbs Energy diagram
Tie lines' calculation	9000	phases composition	Newton	$10^{-15} – 10^{-20}$	OF_2	Common Tangent Plane condition
outer level optimization	300 – 600	both	genetic	10^{-12}	OF_3	Best fitness improvement

Table 2: main tolerances used for the nested optimization algorithms.

Both OF_1 and OF_2 are calculated by an activity minimization procedure. In the first case, the Nelder-Mead algorithm (based on the simplex method [22]) was applied, fixing the phases composition (taken as their values in the genome) and perturbing the thermodynamic parameters. This step is meant to work as a local optimization of the binary parameters around the native values of the individual, and the OF_1 value is the equal to the minimization's output. To calculate OF_2, a modified version of OF_1 is first executed, i.e.: $OF_1{}'(\boldsymbol{x}) = \sum_{i=1}^{3}(a_{i,d}^{I} - a_{i,d}^{II})^2$ at fixed parameters $\underline{U}$, then the best $\boldsymbol{x}$ values are used to calculate OF_2 (which is actually minimized by the outer shell). In this case, a Newton-type method was used [23], after many fruitless trials with the Nelder-Mead one [24], with two different 'receipts' and closure conditions:

a. 'Lever-rule receipt' [16]: the compositions $x_i{}^I$ in one phase are varied, the other phase is then obtained as: $x_i{}^{II} = z_i - \beta x_i{}^I$ (where β is the split ratio); this method is best applied when the overall mixture composition $\boldsymbol{z}$ before the phase-split is given in the reference data, otherwise one can calculate it imposing $\beta = 0.5$ (or any other reasonable value, depending on the reviewed data) as an extra experimental value.
b. 'Sorensen receipt' [8]: the $x_1{}^I$ value is fixed at its experimental value $\bar{x}_1{}^I$, then one different molar fraction in the same phase and two different ones in the second phase are modified.
c. For each method, the molar balance $1 = \sum x_i{}^{I,II}$ can be handled by the used Newton optimization routine, or can be hard-coded within the OF1' calculation, decreasing the number of varied fractions: this has a non-negligible impact on the overall performance, because most Newton optimization packages treat the bound with projection techniques, thus *increasing* the problem's complexity (from the point of view of the time), while the second approach practically reduces it.

The above provisions could be combined in four ways for each thermodynamic method and mixture examined, yielding a total of 16 cases, nevertheless the possible options were reduced testing both the a) and b) receipt, but keeping the fixed algebraic bound within the OF'_1 calculation, because of its faster execution. The a) receipt was applied in the full procedure, while the b) receipt was also tested skipping the optional OF_1 minimization; so twelve cases were set-up.

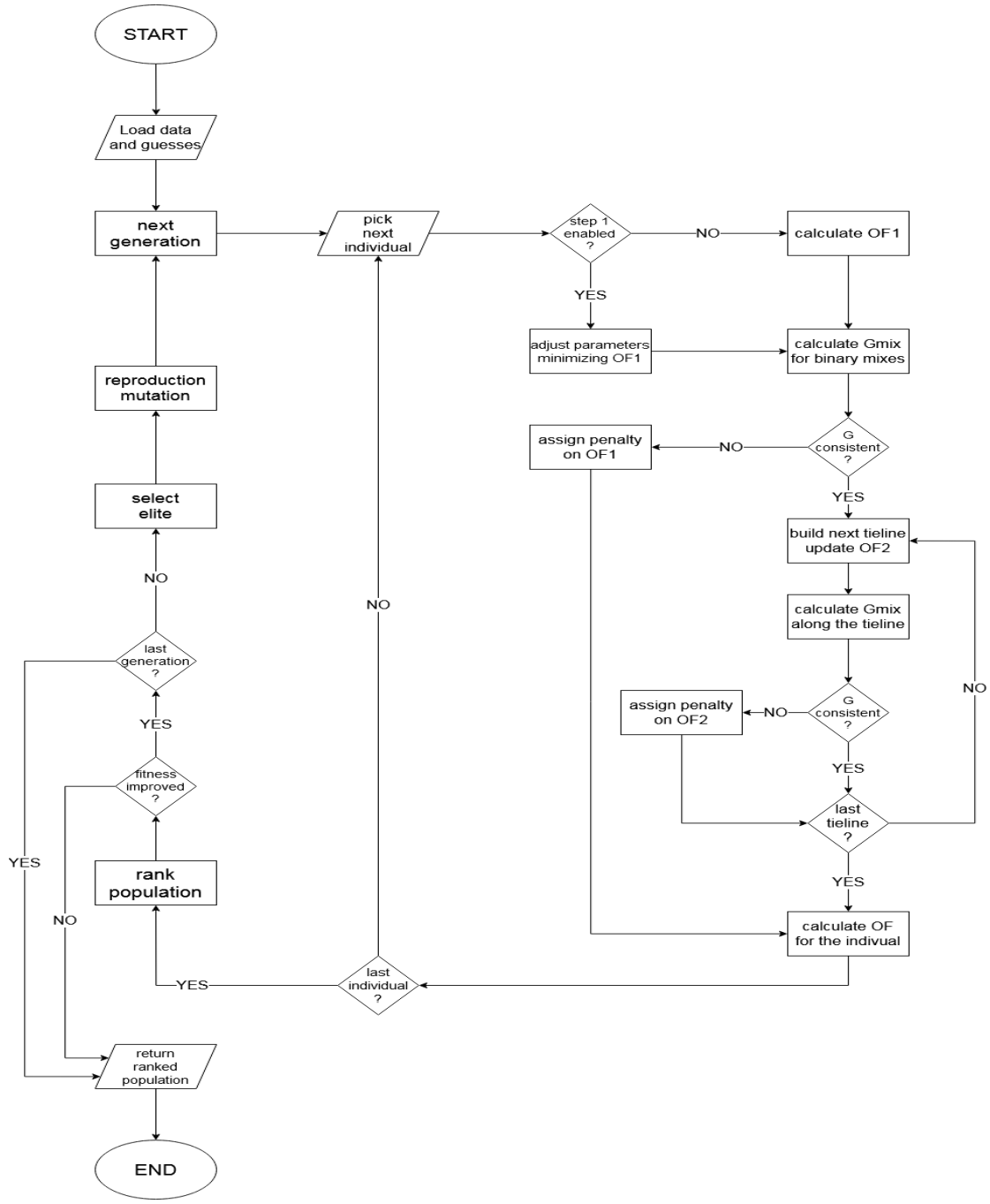

Figure 1: flow diagram of the calculation steps; adapted from the proposal of [16].

Concerning the "outer shell" algorithm (see also Supporting Information), the reproductive function was set to work with the full set of possibilities (see also Figure *2*): a) elite individuals, b) crossover children and c) mutated individuals. The number of crossover over total (crossover plus mutated) children is tuned by a dedicated parameter [25], while a child's gene is derived picking the same gene from one of the parents (at random). The genes mutation for the k^{th} generation consists in the application of a Gaussian perturbation, whose mean is the gene value at the k^{th}*-1* generation, and whose standard deviation is based on the dispersion of that gene tuned by a constant *k'*: $\sigma_k = \sigma_{k-1}\left(1 - \frac{k\prime}{k}\right)$ [26,27].

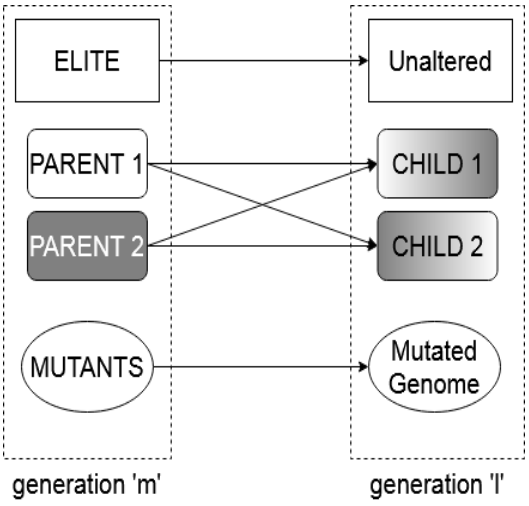

Figure 2: graphical scheme of the reproductive strategies.

Each n^{th} individual $\underline{P}$ is an array of g real values (where $s=3$ are the chemical species and d the experimentally available tie-lines for each system) arranged into a $n \times g$ matrix $\underline{\underline{P}}$:

$$\underline{P} = \left[U_{1,1} \dots U_{s,1}, \dots, U_{1,s} \dots U_{s,s}, \bar{x}^{I}_{1,1} \dots \bar{x}^{I}_{d,1}, \bar{x}^{I}_{1,2} \dots \bar{x}^{I}_{d,2}, \bar{x}^{II}_{1,1} \dots \bar{x}^{II}_{d,1}, \bar{x}^{II}_{1,2} \dots \bar{x}^{II}_{d,2}\right], \qquad \underline{\underline{P}} = \begin{bmatrix} \underline{P}_1 \\ \vdots \\ \underline{P}_n \end{bmatrix}$$

Being the parameter matrix $\underline{\underline{U}}$ actually split into two 3×3 matrixes (see also supporting information), having 6 to 10 experiments (mixture 1 – 2), $g_{mix\,1} = 42$ and $g_{mix\,2} = 58$ (the third molar fraction is not registered because is derived from the other two).

Since the genome of this problem represents two qualitatively different physical quantities (thermodynamic parameters – adjusted phases composition), only the above described method for children's gene is reasonable as, for example, the twentieth gene (a molar fraction) cannot assume the value of an interaction parameter (generally higher than the unity and possibly negative).

The individuals fit for reproduction are ranked according to *their position* in a fitness-based list. The other possible option to rank them on the basis of the *fitness values themselves* was discarded, in view of keeping some "bad" individual as eligible to pass its genes down to the next generation and thus avoiding local minima. When a child is to be spawned, two parents are picked from the list with a probability following the (normalized) cumulative value of the ranks.

Unlike in the reference work [16], where a "$\mu + \lambda$" reproduction strategy is adopted (being μ the individuals going untouched, and λ the freshly spawned ones), a "$\mu, elites + \lambda$" approach was adopted [28]. In this way, being the number of elite individuals small, while adopted μ values are usually larger (see [29] and references therein), the diversity between two generations is increased, while keeping trace of the best results so far [30].

All the above described calculations are performed with a custom-written bundle of Mathwork's Matlab scripts, except for the genetic algorithm (that relies on the Mathwork version of the 'ga' function provided

with the r2020 software), the Nelder-Mead algorithm ('fminsearch', 'fminsearchbnd') and the Newton-type method ('fmincon') [31].

Results

To evaluate the effectiveness of the regression and the quality of the results various indicators have been registered (the bolded entries are listed in Table 3):

1. **number of generations;**
2. best fitness value for each generation;
3. best individual for each generation;
4. **final error on the calculated tie lines and total error;**
5. obtained parameters for the gamma model;
6. **shape of the free energy of Gibbs along the tie line, and CTP fulfillment;**
7. shape of the free energy of Gibbs for each binary mixture.

The first four indicators are useful to evaluate the quality of the data representation and the capability of the algorithm to reach it. The sixth and seventh indicators help to evaluate the thermodynamic consistence of the obtained results.

The first results obtained were used to find out the best combination of algorithm parameters, using accepted literature parameters, before moving on to their refinement. This had an impact on the outlined methodology, since it was found out that the double consistency check, when applied rigorously, could actually prevent any solution to be reached. In particular, three alternative shapes of the ΔG_{mix} function for the binary systems were allowed, increasing the OF_1 value with properly tuned penalties (from the highest):

- local 'minima' with non-zero derivative located at 0 and 1 and one maximum between them;
- one minimum, one maximum and a non-zero-derivative minimum (at 0 or 1);
- more than one maximum, two minima and one non-zero-derivative minimum.

The first two relaxations are needed to account for the space-discretization error involved in any numerical representation, that can "average out" local minima too close to the compositions range extremes (as in the case of water and toluene). The third allowance speeds up the calculation, breaking long tie searches and "saving" some low-scoring individual for the genetic search. Other details are in the supporting info.

Heuristically, it was observed that each time (for every case) the genetic algorithm worked for nearly 50 generations as a "global" minimum optimizer, with large values ($10^1 – 10^2$) of OF_3 due to the penalties imposed on OF_1 and OF_2. After the appearance of an individual with consistent thermodynamic parameters (yielding correct shapes of the free Gibbs energy curves of all the 3 binary mixtures, and respecting the CTP condition for every tie-line), usually towards the $60^{th} – 70^{th}$ generation and with $OF_3 < 1$, the genetic procedure spent all the remaining time to find a refinement of the aforementioned solution.

This behavior repeats also increasing the random disturbances foreseen in the reproduction phase. Set of starting parameters yielding OF_1, OF_2 with penalties results in an early algorithm stop, as the fitness value remains fixed for too many generations: this means that there are regions of the phase space that the algorithm cannot connect to the zone containing the minimum. From this point of view, the choice of the parameters bounds is very important, because they are used to set-up the initial population unless a specific ensemble is manually provided.

The 4 best regression cases (for each thermodynamic method and each mixture) are reported synthetically in Figure 3-Figure 4-Figure 5-Figure 6. The best fits were always obtained searching the tie-lines with the Sorensen receipt. The second mixture was treated satisfactorily with both the UNIQUAC and the NRTL method, while the latter performed poorly with the first mixture: this may be due to the fact of having fixed the non-randomness matrix – nevertheless, there was no need to free the corresponding r and q constants for the UNIQUAC equation.

The evaluation of the critical points was made a posteriori on the calculated data: for the calculation, the Merchuk [32], the conjugation [33] and Othmer-Tobias [34] methods were used, so in order to have an homogenous comparison, the same interpolation were applied also to the literature data, even if in this case the plait-points are already reported by the authors. In general, the conjugation method showed a systematic bias, while the first component fraction is always calculated correctly by the O-T procedure except for Case 9. The second component value is more scattered, for two reasons: the Merchuk procedure is very sensitive to the shape of the binodal derivative, the O-T method instead lacks the data to interpolate accurately the binodal just near the critical point. Nevertheless, at least one of the estimated plait points is close to the literature one, save for the generally bad outcome of case 5. The details can be found in the supporting information.

Case	**Number of cycles / total generations**	**Error on tie lines (OF2)**	**Total error** $\sqrt[2]{\mathbf{OF1 + OF2}}$	**CTP max error / on tie line**
1	1 / 104	2.3×10^{-3}	5.9×10^{-2}	9.69×10^{-7} / 1
2 *	5 / 666	6.7×10^{-4}	3.4×10^{-2}	2.03×10^{-6} / 6
3	1 / 60	2.0×10^{-3}	5.1×10^{-2}	1.13×10^{-6} / 4
4	2 / 381	6.3×10^{-2}	2.5×10^{-1}	3.95×10^{-4} / 4
5 *	1 / 481	3.5×10^{-3}	9.1×10^{-2}	6.09×10^{-3} / 1
6	2 / 571	4.0×10^{-3}	9.6×10^{-2}	6.09×10^{-3} / 1
7	1 / 441	2.3×10^{-1}	8.9×10^{-1}	1.93×10^{-2} / 6
8	2 / 176	9.3×10^{-2}	2.4×10^{0}	6.88×10^{-4} / 1
9 *	1 / 600	4.6×10^{-3}	2.2×10^{0}	1.07×10^{-4} / 2
10	2 / 261	6.6×10^{-2}	1.4×10^{0}	1.90×10^{-3} / 1
11 *	-	1.9×10^{-1}	1.4×10^{0}	9.99×10^{-4} / 3

12	4 / 1600	5.2×10^{-1}	8.0×10^{-1}	3.00×10^{-3} / 1

Table 3: Overall algorithm performance indicators; best results are marked with *.

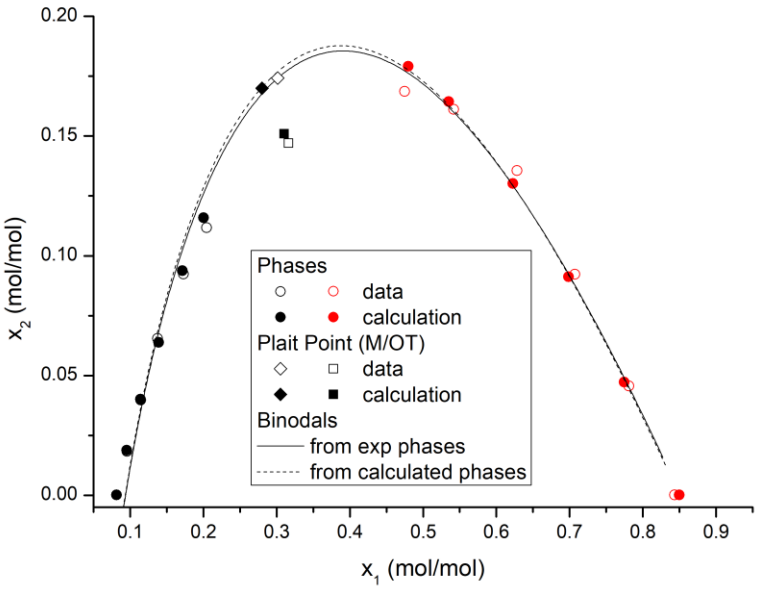

Figure 3: Case 2 (mixture 1, UNIQUAC) phases, plait points and interpolated binodal.

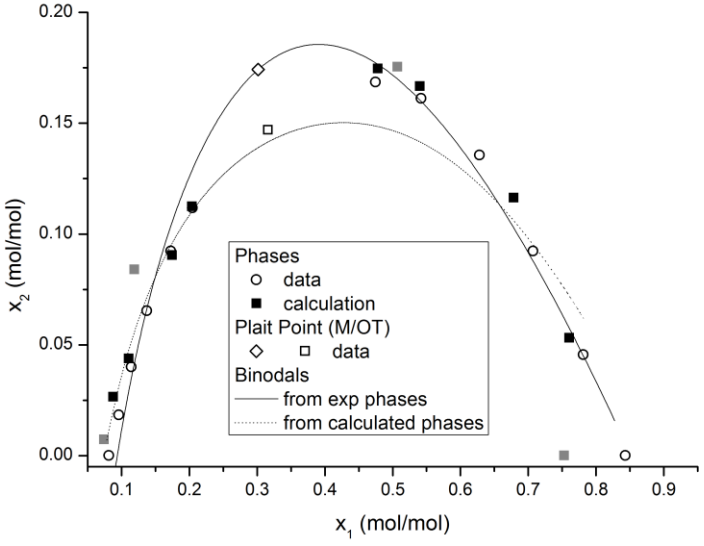

Figure 4: Case 5 (mixture 1, NRTL) phases, plait points and interpolated binodal. Grayed points are unaccepted results.

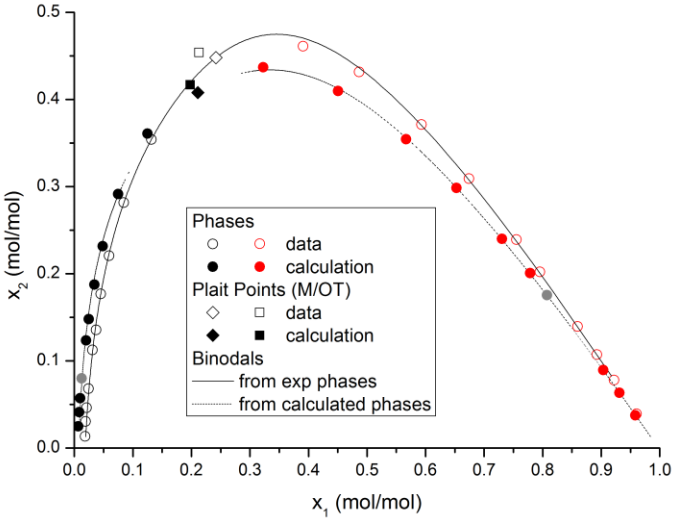

Figure 5: Case 9 (mixture 2, UNIQUAC) phases, plait points and interpolated binodal. Points in grey denote a tie-line with a wrong orientation.

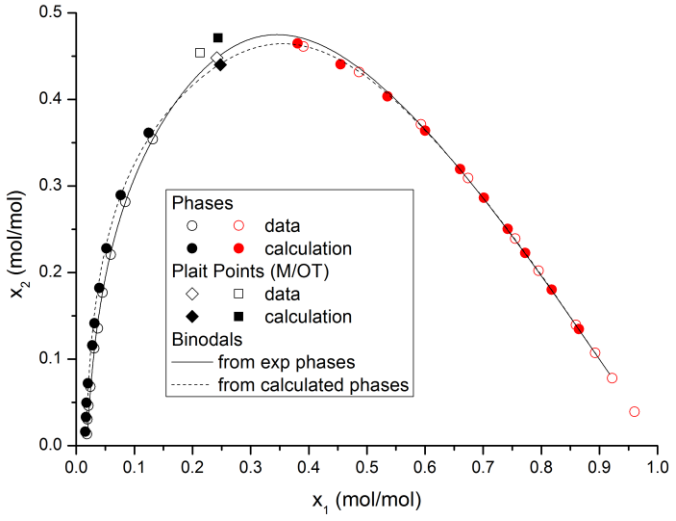

Figure 6: Case 11 (mixture 2, NRTL) phases, plait points and interpolated binodal.

The better performance obtained fixing one component in one phase, with respect to the method of fixing the overall composition split and the lever-rule, is likely due to the fact that, in the second case, a relatively little variation of the b ratio causes higher relative variations in the second phase composition. According to the first receipt, instead, similar relative perturbations of the phases values generate lower relative variation in the estimated b, which eventually reflects in lower values of the residuals.

This issue is shown by examining each tie-line reproduction, not by general features such as the binodal shape or critical point: for example, the NRTL method with the Tomassetti's receipt (case 10) reproduces the experimental binodal region as precisely as using the Sorensen method (case 11 – Figure *7*), and the two cases

yields similar consistency trends and critical points in the Othmer-Tobias test. Nonetheless the latter procedure grants that all the tie-lines slopes and position follow the experimental trend, which is more important than minor deviations in the compositions values. The different behavior of the two models can be also shown by the energy profiles along coexistence lines (Figure *8*).

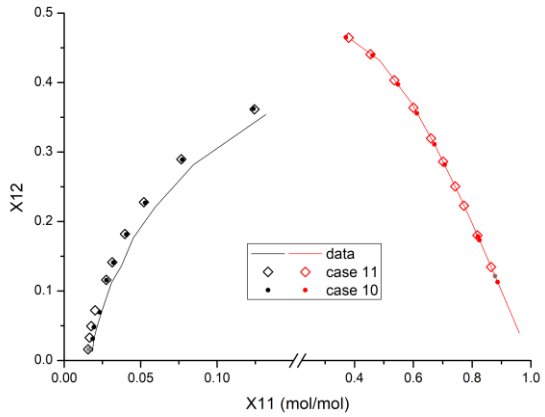

Figure 7: Phases calculated with UNIQUAC-NRTL models (cases 10-11), applied to mixture 2: though case 10 binodal is slightly closer to the experimental one, the tie-lines alignment is worst.

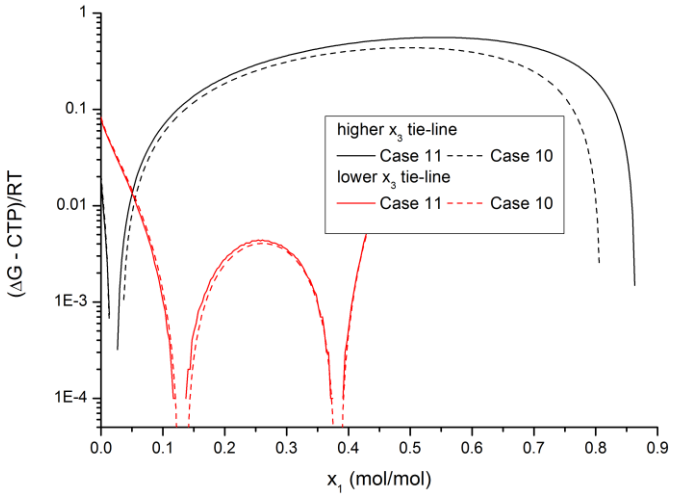

Figure 8: Gibb's energy along the common tangent line direction for the widest (high x3 fraction) and shortest (low x3 fraction) tie-line of mixture 2, calculated with UNIQUAC (case 10) and NRTL (case 11).

Figure 9 and Figure 10 present the excess free energy calculated for the binary mixture that show a miscibility gap (and that determines the ternary equilibria reviewed), comparing the best UNIQUAC (case 2 – 9) and NRTL (cases 5 – 11) for the heptane-methanol and water-toluene couples respectively. It can be seen that both the experimental energy minima are correctly located in every case, while it is the general shape of the NRTL-calculated energy for heptane and methanol (without an appreciable barrier between the minima) to determine also the poor representation of the heptane-benzene-methanol system far from the critical point.

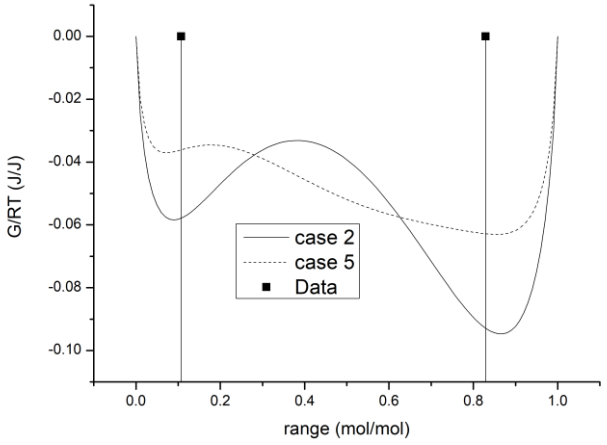

Figure 9: UNIQUAC (case 2) and NRTL (case 5) Gibbs energy for the binary mixture heptane-methanol. The coexisting phases experimental compositions are that of Katayama et al. [35].

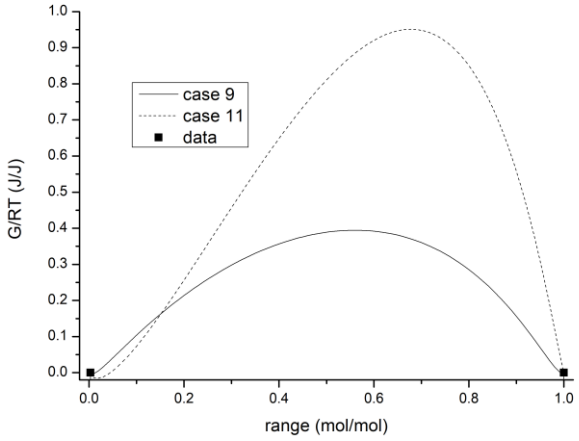

Figure 10: UNIQUAC (case 9) and NRTL (case 11) Gibbs energy for the binary mixture water-toluene. The coexisting phases experimental compositions are that of Neely et al. [36].

Conclusions

Even though mixture 1 was originally presented by Anderson and Prausnitz as an example to extend the UNIQUAC model (considering implicitly a ternary interaction contribution), this work confirms the conclusion that an efficient regression method can solve the problem with the basic UNIQUAC equation, and extends this framework capability to the NRTL formula. It was found heuristically that the differences in the two models have an impact on the tie-lines calculation, while the outer genetic algorithm remained unaffected,

as expected. This implies that the techniques for the regression of binary interaction parameters deserves by themselves further attention, as the full potential of the basic UNIQUAC/NRTL models could still be exploited, without increasing the number of parameters or resorting to equations refinements. Also the choice of fixing the NRTL non-randomness matrix for this calculation is done in this view.

The Nelder-Mead (simplex) algorithm, working on the isoactivity condition only, never yielded a good reproduction of the tie-lines: for this step, only the Newton-type methods could be used successfully. Probably this is due to activity minima following qualitatively the excess energy minima, which makes a gradient-based methodology preferable.

It was observed that the (optional) first step parameters modification improves the algorithm performance when using UNIQUAC, but not when using NRTL. This may imply that, using UNIQUAC, a phase composition is more important to determine activity values than variations in the parameters, while with NRTL a modification in the parameter requires appreciable composition adjustments in the following step. So, in the first case the genome perturbation just makes the algorithm to explore wider phase-space zones, while in the second case the tie-line reproduction is often jeopardized.

The algorithm can deal with different binodal amplitudes, and the hybrid penalties strategy used to assess the ΔG_{mix} correctness succeeds in yielding consistent parameters, even with functions having a different number of minima and flexes with respect to the ideal case.

This work is still open to several improvements: direct minimization of the dGmix residual respect to the tangent plane and adaptive space discretization near dG minima and domain extremes (in the inner level calculations), parallel treatment of VLE and other reproduction function options at the outer level.

Bibliography

[1] H. R, J.M. P. Local Compositions in Thermodynamic Excess Functions for Liquid Mixtures. AICHE J 1968;14:135–44.

[2] Abrams DS, Prausnitz JM. Statistical thermodynamics of liquid mixtures: A new expression for the excess Gibbs energy of partly or completely miscible systems. AIChE J 1975;21:116–28. https://doi.org/10.1002/aic.690210115.

[3] Prausnitz JM, Tavares FW. Thermodynamics of Fluid-Phase Equilibria for Standard Chemical Engineering Operations. AIChE J 2004;50:739–61. https://doi.org/10.1002/aic.10069.

[4] O'Connell JP, Gani R, Mathias PM, Maurer G, Olson JD, Crafts PA. Thermodynamic property modeling for chemical process and product engineering: Some perspectives. Ind Eng Chem Res 2009;48:4619–37. https://doi.org/10.1021/ie801535a.

[5] J. M. Prausnitz, C. A. Eckert, R. V. Orye and JPO. Computer Calculations for Multicomponent Vapor-

Liquid Equilibria. Englewood Cliffs, New Jersey: Prentice-Hall; 1967.

[6] Marcilla A, Olaya MM. Should we trust all the published LLE correlation parameters in phase equilibria? Necessity of their assessment prior to publication. Fluid Phase Equilib 2016. https://doi.org/10.1016/j.fluid.2016.11.009.

[7] Marcilla Gomis A. GE Models and Algorithms for Condensed Phase Equilibrium Data Regression in Ternary Systems: Limitations and Proposals. Open Thermodyn J 2011;5:48–62. https://doi.org/10.2174/1874396x01105010048.

[8] Sørensen JM, Magnussen T, Rasmussen P, Fredenslund A. Liquid-liquid equilibrium data: Their retrieval, correlation and prediction Part II: Correlation. Fluid Phase Equilib 1979;3:47–82. https://doi.org/10.1016/0378-3812(79)80027-8.

[9] Marcilla A, Olaya MM, Reyes-Labarta JA. Simultaneous VLLE data correlation for ternary systems: Modification of the NRTL equation for improved calculations. Fluid Phase Equilib 2016;426:47–55. https://doi.org/10.1016/j.fluid.2015.12.047.

[10] Bollas GM, Barton PI, Mitsos A. Bilevel optimization formulation for parameter estimation in vapor-liquid(-liquid) phase equilibrium problems. Chem Eng Sci 2009;64:1768–83. https://doi.org/10.1016/j.ces.2009.01.003.

[11] Zhu Y, Xu Z. Calculation of liquid-liquid equilibrium based on the global stability analysis for ternary mixtures by using a novel branch and bound algorithm: Application to UNIQUAC equation. Ind Eng Chem Res 1999;38:3549–56. https://doi.org/10.1021/ie990104m.

[12] Tessier SR, Brennecke JF, Stadtherr MA. Reliable phase stability analysis for excess Gibbs energy models. Chem Eng Sci 2000;55:1785–96. https://doi.org/10.1016/S0009-2509(99)00442-X.

[13] Díaz I, Rodríguez M, González EJ, González-Miquel M. A simple and reliable procedure to accurately estimate NRTL interaction parameters from liquid-liquid equilibrium data. Chem Eng Sci 2019;193:370–8. https://doi.org/10.1016/j.ces.2018.09.015.

[14] Anderson TF, Prausnitz JM. Application of the UNIQUAC Equation to Calculation of Multicomponent Phase Equilibria. 2. Liquid-Liquid Equilibria. Ind Eng Chem Process Des Dev 1978;17:552–61. https://doi.org/10.1021/i260068a029.

[15] Santori G, Franciolini M, Di Nicola G, Polonara F, Brandani S, Stryjek R. An algorithm for the regression of the UNIQUAC interaction parameters in liquid-liquid equilibrium for single- and multi-temperature experimental data. Fluid Phase Equilib 2014;374:79–85. https://doi.org/10.1016/j.fluid.2014.04.014.

[16] Tomassetti S, Di Nicola G, Santori G. Identification of UNIQUAC binary interaction parameters in liquid-liquid equilibrium. Fluid Phase Equilib 2020;510:112483.

https://doi.org/10.1016/j.fluid.2020.112483.

[17] SCHWEFEL H-GB and H-P. Evolution strategies: A comprehensive introduction HANS-GEORG. Nat Comput 2002;1:3–52. https://doi.org/10.1023/A:1015059928466.

[18] Beyer HG. An alternative explanation for the manner in which genetic algorithms operate. BioSystems 1997;41:1–15. https://doi.org/10.1016/S0303-2647(96)01657-7.

[19] Gibbs JW. a Method of Geometrical Representation of the Thermodynamic Properties of Substances By Means of Surfaces. Trans Connect Acad 1873;II:382–404.

[20] García-Flores BE, Águila-Hernández J, García-Sánchez F, Aquino-Olivos MA. (Liquid–liquid) equilibria for ternary and quaternary systems of representative compounds of gasoline + methanol at 293.15 K: Experimental data and correlation. Fluid Phase Equilib 2013;348:60–9. https://doi.org/10.1016/j.fluid.2013.03.022.

[21] Gramajo De Doz MB, Bonatti CM, Sólimo HN. Liquid-liquid equilibria of ternary and quaternary systems with two hydrocarbons, an alcohol, and water at 303.15 K: Systems containing 2,2,4-trimethylpentane, toluene, methanol, and water, or 2,2,4-trimethylpentane, toluene, ethanol, and water. Fluid Phase Equilib 2003;205:53–67. https://doi.org/10.1016/S0378-3812(02)00268-6.

[22] Nelder JA, Mead R. A Simplex Method for Function Minimization. Comput J 1965;7:308–13. https://doi.org/10.1093/comjnl/7.4.308.

[23] Coleman T, Li Y. On the Convergence of Reflective Newton Methods for Large-scale Nonlinear Minimization Subject to Bounds vol. 67. Ithaca, NY, USA Cornell Univ 1994.

[24] Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim 1998;9:112–47. https://doi.org/10.1137/S1052623496303470.

[25] Dang DC, Friedrich T, Kötzing T, Krejca MS, Lehre PK, Oliveto PS, et al. Escaping Local Optima Using Crossover with Emergent Diversity. IEEE Trans Evol Comput 2018;22:484–97. https://doi.org/10.1109/TEVC.2017.2724201.

[26] Murata T, Ishibuchi H. Positive and negative combination effects of crossover and mutation operators in sequencing problems 1996:170–5.

[27] Ye F, Wang H, Doerr C, Bäck T. Benchmarking a (u+l) Genetic Algorithm with Configurable Crossover Probability. Lect Notes Comput Sci 2020;12270 LNCS:699–713. https://doi.org/10.1007/978-3-030-58115-2_49.

[28] HANS-GEORG BEYER. The Theory of Evolution Strategies. Berlin Heidelberg: Springer; 2001. https://doi.org/10.1007/978-3-662-04378-3.

[29] Back T, Hammel U, Schwefel HP. Evolutionary computation: Comments on the history and current state. IEEE Trans Evol Comput 1997;1:3–17. https://doi.org/10.1109/4235.585888.

[30] Doerr B, Doerr C, Ebel F. Lessons from the black-box: Fast crossover-based genetic algorithms. GECCO 2013 - Proc 2013 Genet Evol Comput Conf 2013:781–8. https://doi.org/10.1145/2463372.2463480.

[31] MATLAB. (R2020b). Natick, Massachusetts: The MathWorks Inc.; 2020.

[32] Merchuk JC, Andrews BA, Asenjo JA. Aqueous two-phase systems for protein separation. J Chromatogr B 1998;711:285–93. https://doi.org/10.1016/S0378-4347(97)00594-X.

[33] Weber HC, Meissner P. Thermodynamics for chemical engineers. II. New York: John Wiley & Sons; 1957.

[34] Othmer DF, Tobias E. TIE LINE CORRELATION. Ind Eng Chem 1942;34:693–6.

[35] Katayama H, Ichikawa M aki. Liquid-liquid equilibria of three ternary systems: Methanol-heptane including 1,3-dioxolane, 1,4-dioxane and tetrahydropyran in the range of 253.15 to 303.15k. J Chem Eng Japan 1995;28:412–8. https://doi.org/10.1252/jcej.28.412.

[36] Neely BJ, Wagner J, Robinson RL, Gasem KAM. Mutual solubility measurements of hydrocarbon-water systems containing benzene, toluene, and 3-methylpentane. J Chem Eng Data 2008;53:165–74. https://doi.org/10.1021/je700449z.