

Techniques for assessing the effects of uncertainties in thermodynamic models and data

W.B. Whiting^{a,*}, V.R. Vasquez^a, M.M. Meerschaert^b

^a *Chemical and Metallurgical Engineering Department, University of Nevada, Reno, Reno, NV 89557-0136, USA*

^b *Mathematics Department, University of Nevada, Reno, Reno, NV 89557-0136, USA*

Received 11 April 1998; accepted 25 January 1999

Abstract

Thermodynamic models and experimental data exhibit the usual systematic and random errors. The severity of their errors depends on their use, such as for process calculations in a process simulator. Similarly, the value of better thermodynamic models and/or data should be measured with reference to such use. We have developed techniques for quantification of such thermodynamic-induced process uncertainties via Monte Carlo simulation, regression analyses, and analogies to optimization. The influence of experimental data sources and data types on the uncertainty of thermodynamic models is studied. Details and applications of our new sampling strategy (EPS), which accounts for the high degree of correlation between thermodynamics model parameters, is given. This procedure directly uses the regression results in a way that is much more powerful and mathematically accurate than traditional covariance matrix techniques. Level sets are used for the Monte Carlo samples so that unbiased accurate sampling of the entire feasible region is obtained. Comparison with traditional Monte Carlo sampling, Latin Hypercube sampling (LHS), and Shifted Hammersley sampling (SHS) are shown. The result is an unbiased estimate of uncertainties that reduces the over- and under-estimations common in traditional techniques. The approaches presented can be used for safety-factor/risk analysis, guidelines for simulator use, experimental design, and model comparisons. They allow determinations of the value of obtaining additional phase-equilibrium data and the potential value of improved phase-equilibrium models. Examples and case studies of these applications are provided. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Sensitivity; Uncertainty; Method of calculation; Liquid–liquid equilibria; Models; Sampling techniques

* Corresponding author. Tel.: +1-775-784-4307; Fax: +1-775-784-4764; E-mail: wwhting@unr.edu

1. Introduction

The effects of uncertainties in thermodynamic models and data play an important role in predicting process performance, simulation, process design, and optimization. Techniques for assessing these effects are required for the continuous improvement of thermodynamic models and for the design of experiments. During the last 10 years, uncertainty analysis in the chemical industry has been focused on areas like process synthesis under uncertain conditions (Diwekar and Rubin [1], Chaudhuri and Diwekar [2]), process design (Halemane and Grossman [3], Pistikopoulos and Ierapetritou [4]), stochastic planning (Ierapetritou et al. [5]) and so on. Surprisingly, there are few studies reported in the literature on the analysis of uncertainty associated with thermodynamic models, which play a very important role in process design and simulation.

The effect of property inaccuracies on process design is also of great importance in the chemical industry. Awareness of this problem has been reported earlier by Streich and Kistenmacher [6], Nelson et al. [7], and Macchietto et al. [8]. More recent studies confirm the existence of this problem by analyzing the effect of thermodynamic data on estimated process performance (Reed and Whiting [9]; Whiting et al. [10]; Whiting [11]). In general, thermodynamic models are greatly affected by property inaccuracies because most of the parameters in the models are obtained via nonlinear regression procedures using experimental data.

In this work, techniques for quantification of thermodynamic-induced process uncertainties are presented via Monte Carlo simulation coupled with regression analysis. In this method, we try to understand the uncertainty in the final calculations by doing the calculations many times, each time using different values of the model parameters. We choose the model parameter values statistically, using our knowledge of their uncertainties, which are a direct result of the regression of the experimental data. Thus, we simulate on a computer what might logically occur if we were to perform the experiments many times, each time regressing to find the best values of the model parameters.

The first issue addressed is the effect that the source of experimental data has on parameter regression for thermodynamic models, and how this effect is propagated into uncertainty of predicted process performance. The results show that the effect is significant, and the method proposed for quantifying it also allows the detection of underdesign and overdesign problems, which are hidden under classical design methods. Secondly, the effect of data type on parameters regression is studied. For instance, to obtain the binary interaction parameters of a model like NRTL or UNIQUAC for a ternary liquid–liquid Type II system, binary liquid–liquid, ternary liquid–liquid, and binary vapor–liquid data can be used. The effect of the data type is very significant on the sensitivity and uncertainty analysis of chemical processes. The last issue addressed in this work is the effect of using different sampling techniques to perform the uncertainty analysis.

The mathematical complexities of thermodynamic models, which are highly nonlinear, present problems for the commonly used Monte Carlo sampling techniques (such as Latin Hypercube sampling (LHS) and Shifted Hammersley sampling (SHS)). These traditional techniques are analogous to first-order experimental design strategies, in which only first-order corrections are made for correlations between model parameters. A new sampling strategy (EPS), which accounts for the high degree of correlation between thermodynamic model parameters is proposed. Better sampling of the parameter space is obtained than that obtained from the commonly used techniques, which are based on linear approximations using the covariance matrix. All the methods and techniques suggested are analyzed through the use of case studies.

The approaches presented can be used for safety-factor/risk analysis, guidelines for simulator use, experimental design, model comparisons, and model regression. They allow evaluation of the value of obtaining additional phase-equilibrium data and the potential of improved phase-equilibrium models.

2. Uncertainty of thermodynamic models and data

The uncertainty associated with thermodynamic models can be described in terms of the model parameter uncertainty and modeling error. There are two main sources of uncertainty for parameter values: (1) the method used for obtaining the parameters, and (2) the experimental data. The modeling error deals with the parameterization chosen to describe the physical system under study, which always is an approximation of the real system.

The method used for obtaining the parameters in a given thermodynamic model depends on the definition of the objective function, which can have different forms (e.g., see Sørensen et al. [12], Esposito and Floudas [13], Reid et al. [14]) and on the numerical technique used for minimizing or maximizing the objective function. The effect of regression methods on the model parameter uncertainties is an active area of optimization research. Of special concern are local minima in the Gibbs free energy surface. If there is significant modeling error, and/or systematic trends in the data, local minima can grow and the solution can lead to spurious phases (Sørensen et al. [12]).

The parameterization chosen affects substantially the description of the chemical system. For example, Fig. 1 presents the binodal curve estimation for the ternary system 1,1,2-trichloroethane + acetic acid + water at 25°C using two activity coefficient models for liquid–liquid equilibria prediction (UNIQUAC and NRTL). The significance of the modeling error is substantial. However, the following question arises: how significant is this error in process design and simulation? In order

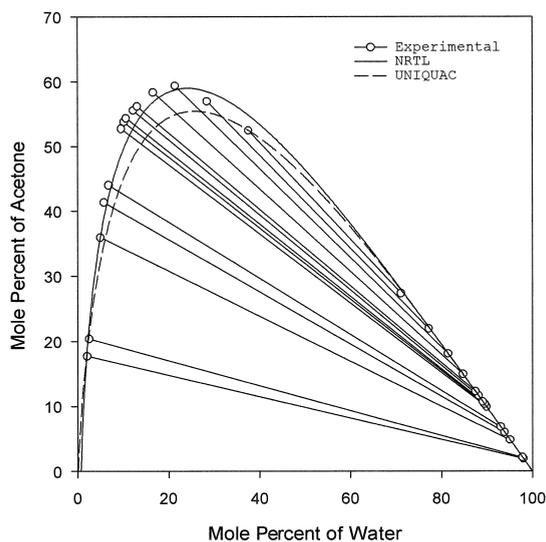


Fig. 1. Liquid–liquid equilibria for 1,1,2-trichloroethane + acetone + water system at 25°C. The predicted values are by the UNIQUAC and NRTL models. Experimental data are from Sørensen and Arlt [19].

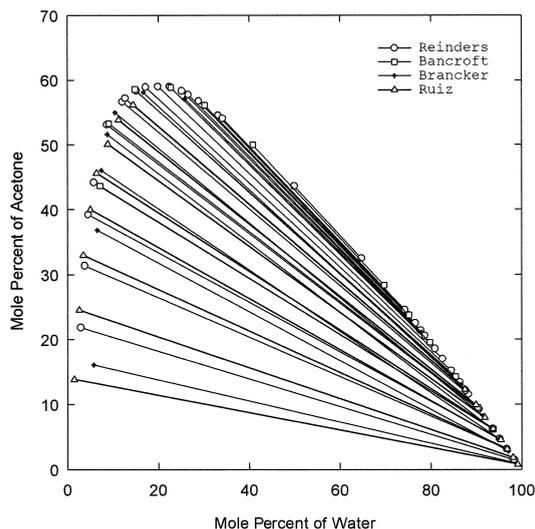


Fig. 2. Liquid–liquid equilibria for the chloroform + acetone + water ternary system at 25°C from different experimental data sources.

to answer this question, other factors have to be taken into account: the effect of the experimental data uncertainty on the parameter fitting process, and the effect of the experimental data uncertainty on the generated binodal curves.

The uncertainty associated with experimental data is usually described in terms of random and systematic errors. Often, the random errors are assumed to reduce to the *apparatus precision*, which would not cause major impact on the parameter uncertainty for a given model. The systematic uncertainties are conceptually well defined, but they are traditionally ignored in regression. Some authors (e.g., Dietrich [16]) consider the effect of systematic errors as a shift of the mean by a given value from many repeated experiments, and simple rectangular probability distributions are used. Even though for some applications this may be a good definition of this kind of uncertainty, the problem is more complex for thermodynamic data. Fig. 2 shows the liquid–liquid equilibria for the chloroform + acetone + water ternary system at 25°C from different experimental data sources. The presence of systematic trends in the data is evident, but the shift is basically on the left phase only.

3. Assessing effects of uncertainties

Assessing effects of uncertainties in thermodynamic models can be carried out in several ways depending on the nature of the problem being analyzed. In this work, a computer-based approach through stochastic models and simulation is used. The approach is summarized as follows:

1. specification of probability distributions for uncertain input variables, such as model parameters;
2. sampling from these probability distributions to mimic repeated experimental runs;
3. simulation of the process or phase equilibrium with each set of model parameters;
4. statistical analysis of results to provide quantitative uncertainty measures for the system.

Details of the application of these steps are explained in the following sections.

3.1. Data source effect on uncertainty of thermodynamic models

The effect of the experimental data source on the uncertainty of thermodynamic models is studied using different data sets for the regression of NRTL binary interaction parameters (for details of the NRTL equation see Novák et al. [17]). The data sets are taken from the literature, and regressions are performed to obtain the parameters b_{ij} and b_{ji} , fixing α_{ij} at 0.2, which is a common simplification.

The regression method used consists of minimizing the following objective function for the differences between measured data and the model prediction using the maximum likelihood principle:

$$\min S = \sum_i \left[\frac{(T_i^e - T_i^M)^2}{\sigma_{T_i}^2} + \frac{(x_{11_i}^e - x_{11_i}^M)^2}{\sigma_{x_{11_i}}^2} + \frac{(x_{12_i}^e - x_{12_i}^M)^2}{\sigma_{x_{12_i}}^2} \right] + \sum_i \left[\frac{(x_{21_i}^e - x_{21_i}^M)^2}{\sigma_{x_{21_i}}^2} + \frac{(x_{22_i}^e - x_{22_i}^M)^2}{\sigma_{x_{22_i}}^2} \right] \quad (1)$$

where x_{21} is the mole fraction of component 1 in phase 2, and the summations are over all tie lines.

A standard deviation of 0.01°C was used for the temperature and of 0.1% for the compositions (i.e., the standard deviation for a mole fraction of 0.1 is 0.0001) were used for the first two components (the third component is set by difference). The regressions were performed using the ASPEN PLUS (Aspen Tech, Cambridge, MA) simulator. The regression results consist of six regressed parameters, variances and the correlation matrix.

With this standard regression information, a Monte Carlo simulation with simple LHS was used to obtain 100 sets of binary interaction parameters. These 100 sets of parameters represent a statistical sampling of the likely model parameters if, indeed, we could measure absolutely precise and accurate experimental data. These are used to perform 100 simulations on a conventional liquid–liquid extractor, using case studies from the literature. The process simulations were carried out using the ChemCAD (Chemstations, Houston, TX) process simulator.

The ternary liquid–liquid system chloroform(1) + acetone(2) + water(3) is used to illustrate the importance of the data source on the uncertainty analysis of thermodynamic models (for additional systems and case studies see Vasquez and Whiting [18]). Sørensen and Arlt [19] report the experimental data used in this work, in particular the ones originally reported by Brancker et al. [20] (Set 1), Bancroft and Hubard [21] (Set 2), Ruiz and Prats [23], and Reinders and DeMinjer [22] (Set 3) were used to perform the regressions. The experimental tie lines of these systems are presented in Fig. 2. The binary interaction parameters obtained with some statistical measures are presented in Table 1. There are significant differences in the values of the binary interaction parameters and their statistics, depending on the data source, as expected. Fig. 3 presents the three estimated binodal curves (for convenience the experimental tie lines are not presented in this figure. It can be observed that the differences in the predicted binodal curves are significant. However, the engineering significance of these differences can only be determined through actual process calculations.

As an illustrative example, an extraction operation reported by Smith [24] is used. In this operation, water is used to separate a chloroform + acetone mixture in a countercurrent extraction column with two equilibrium stages. The feed consists of equal amounts of chloroform and acetone on a weight

Table 1

Binary parameters b_{ij} and b_{ji} regressed for the NRTL model ($\alpha = 0.2$). System^a: chloroform(1) + acetone(2) + water(3) at 25°C

i	j	b_{ij} (K)	Standard deviation (K)	b_{ji} (K)	Standard deviation (K)
Set 1 (S/n) ^{0.5} = 133.89					
1	2	334.71	985.36	-623.27	519.65
1	3	518.27	214.56	1554.90	174.54
2	3	378.08	554.84	282.94	285.88
Set 2 (S/n) ^{0.5} = 71.77					
1	2	34.30	5.29	-681.75	3.24
1	3	1636.27	36.30	1744.10	2.03
2	3	80.26	1.62	392.63	0.87
Set 3 (S/n) ^{0.5} = 128.25					
1	2	-300.07	273.73	-227.84	278.75
1	3	1044.28	168.30	1499.60	74.07
2	3	-43.37	120.13	518.20	96.08

^aAll the experimental data used for the parameters regressions are from Sørensen and Arlt [19], and the specific references are described in the text.

basis. The operating conditions are 25°C and 1 atm. A mass solvent/feed ratio of 1.565 is used. The output variable is the percentage of acetone extracted. In Fig. 4, the cumulative frequency curves for the percent extracted of the main component are presented, quantifying the uncertainty of the output variables. Such curves are the primary output from Monte Carlo uncertainty analyses, and they

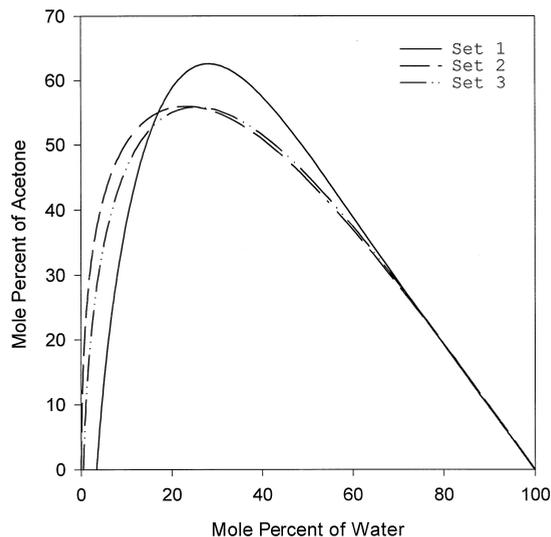


Fig. 3. Predicted binodal curves for the liquid–liquid equilibria of the ternary system chloroform + acetone + water 25°C using the NRTL model. NRTL parameters regressed from different experimental data sources as indicated.

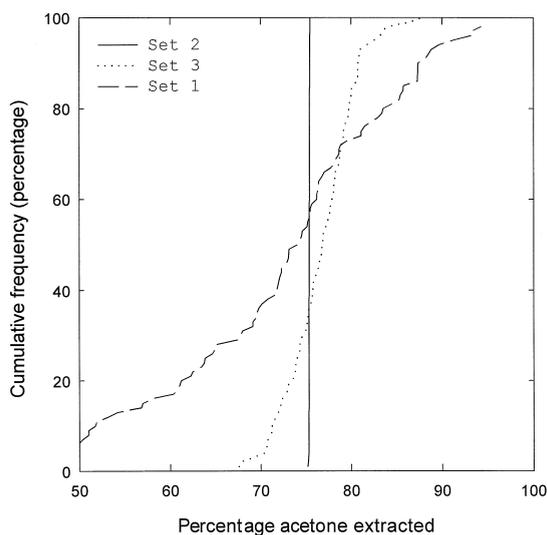


Fig. 4. Uncertainty of percentage of acetone extracted in the liquid–liquid extractor, using different sets of parameters for the NRTL equation.

provide a quantification of risk for the design engineer. For example, we see for the curve labeled ‘Set 1’ that there is a one in five chance that only about 62% of the acetone will be extracted—even though a single design calculation at the best set of model parameters would suggest that we would achieve 75% extraction! Conversely, there is a small chance (about 1 in 10) that we could achieve 88% extraction. This uncertainty in the performance of the extractor arises from the imprecise nature of our knowledge of the model parameters, which in turn arises (in part) from the experimental errors and the fact that the model does not incorporate all of the physics of the system. If we, therefore, wish to provide a guarantee for the performance of the extractor, we use our risk tolerance and the cumulative frequency curve to arrive at a value.

Three of the 100 simulations resulted in unfeasible situations for Set 1, explaining why the cumulative frequency curve does not reach the value of 100 on the vertical axis. Additionally, this distribution is broad compared to the other ones. With this kind of result, clearly care must be taken in selecting the parameters. For more details about these issues and further statistical details, refer to Vasquez and Whiting [18] and Tarantola [15].

This example clearly demonstrates the hidden underdesign and overdesign bias that may result from using the optimum parameters regressed from a given data set. All the data sets analyzed result in underdesign, in which the percentage of acetone extracted is more likely to be less (around 54% of the time) than that calculated with the ‘optimum’ parameters (75.47% for set 1, 75.37% for set 2, and 77.47% for set 3). The true uncertainty is greater than that presented here, because of model weakness and possible existence of systematic errors in the experimental data. Other cases studied (Vasquez and Whiting [18]) show important variations in the median performance even with small amplitude in the distributions, showing that the breadth of the cumulative frequency curve cannot be used as a unique criterion for choosing the best model or parameter set. Also, it can be observed that the difference between uncertainties cannot be inferred from the goodness of fit of the regressions. Additionally, pooling the experimental data sets will not improve the uncertainty estimates, producing only a single

uncertainty analysis. Performing the analysis separately for different data sets gives detailed information for assessing uncertainty and risk.

3.2. Data type effect on parameters regression

The data type effect on parameter regression was studied using two approaches to obtain the binary interaction parameters of the UNIQUAC (Abrams and Prausnitz [25]) model. The first one consists of using binary vapor–liquid and liquid–liquid equilibrium data, and the second one using only liquid–liquid ternary data. Comparisons between these two approaches are done through predicted process performance calculations on a given unit operation. The data sets used are from the literature, and the regressions are for obtaining the parameters b_{ij} and b_{ji} , with any other extra parameters set to zero (i.e., a_{ij} , a_{ji} , c_{ij} , and d_{ij} in the formulation presented by Aspen Technology [26]).

The regression method, objective function, and simulation methodology are the same as explained above. The simulation results are processed similarly. The ternary system chloroform(1) + acetone(2) + water(3) is again used for performing liquid–liquid equilibria computations; however, additional systems and case studies are also given by Vasquez and Whiting [27] for this problem. For the ternary liquid–liquid system, the data reported by Bancroft and Hubard [21] (see experimental tie lines for this system in Fig. 2) are used. The data used for the binary acetone + chloroform vapor–liquid system are reported by Kogan and Deizenrot [28] and Apelblat et al. [29]. The experimental vapor–liquid data for the binary acetone + water were obtained from Eduljee et al. [30], and liquid–liquid data for the binary chloroform + water are from Stephenson [31].

The binary interaction parameters obtained together with the estimated statistics are presented in Table 2. As in the previous section, there are significant differences in the values obtained for the binary interaction parameters and in their statistics. Fig. 5 presents the estimated binodal curves for the three different sets of parameters obtained from regression. The two curves titled ‘Binary_01’ and

Table 2

Binary parameters b_{ij} and b_{ji} regressed for the UNIQUAC model for the system chloroform(1)+acetone(2)+water(3)

i	j	b_{ij} (K)	Standard deviation (K)	b_{ji} (K)	Standard deviation (K)
		TERNARY LLE		$(S/n)^{0.5} = 378.30$	
1	2	241.54	267.32	−94.22	352.29
1	3	−1064.71	941.83	−332.77	37.92
2	3	−129.72	118.15	15.78	34.91
		BINARY Sets		$(S/n)^{0.5}$ (Set 1) = 488.52 $(S/n)^{0.5}$ (Set 2) = 436.38	
1	2	−15.99 ^a	45.40	100.76 ^a	38.70
1	2	−254.28 ^b	540.81	272.70 ^b	249.51
1	3	−813.32	275.31	−342.50	118.50
2	3	−330.45	22.76	35.82	11.45

^aBinary interaction parameter used in Fig. 5 for the ‘Binary_02’ curve. Parameters regressed using data from Apelblat et al. [29].

^bBinary interaction parameter used in Fig. 5 for the ‘Binary_01’ curve. Parameters regressed using data from Kogan and Deizenrot [28].

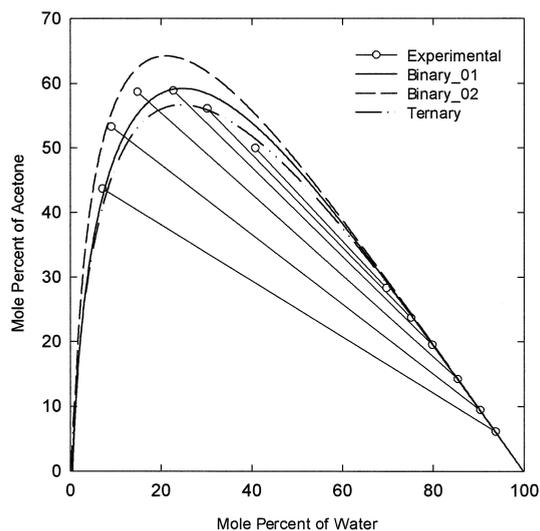


Fig. 5. LLE for chloroform + acetone + water ternary system at 25°C. The predicted binodal curves are from the UNIQUAC model with parameters regressed from binary and ternary data. See text for details of the data sources used.

'Binary_02' represent the difference obtained from using different data sources for the binary system acetone + chloroform keeping the other two binary interaction parameters constant. The estimated equilibria is for the Bancroft and Hubard [21] data.

The effect of using these different sets of binary interaction parameters for a specific chemical process is illustrated through the simulation of a liquid–liquid extraction operation. The case study described in the previous section is used for this purpose. Fig. 6 presents the cumulative frequency

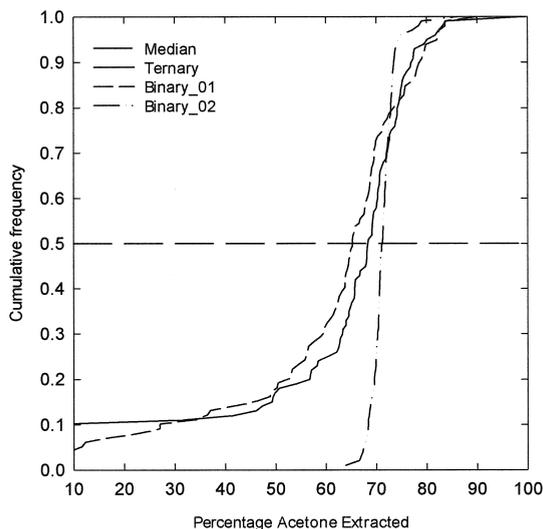


Fig. 6. Uncertainty of percentage of acetone extracted in the liquid–liquid extractor, using different sets of parameters for the UNIQUAC model.

curves for the performance of the extraction simulated. Nine simulations resulted in unfeasible situations for the parameters regressed from the ternary data and one for the other sets. The cumulative frequency curves ‘Binary_01’ and ‘Binary_02’ present very different breadth characteristics. This effect is only due to the difference between the variances of the parameters for the binary acetone + chloroform. The cumulative frequency curve for the parameters regressed from ternary data is broader than the ‘Binary_02’ curve, which is a significant difference for the process performance, and it may increase as the complexity of the process increases.

Thus, the variances of the parameters are very important, but these variances depend on the quality of the experimental data, data type, and model chosen.

3.3. Effect of sampling techniques

The final effect studied is from the sampling technique used. Traditional sampling techniques determine the sets of model parameters use random sampling of the individual model parameters over an assumed probability distribution, often a Gaussian distribution. The regression itself provides a mean and a standard deviation for each parameter, the data needed for the sampling. After an appropriate number of samples are generated, the values of the two (or more) parameters are paired using the correlation matrix (also an outcome of the regression). Such a procedure provides a first-order correction for the correlation between the model parameters, such as is often presented in a ‘confidence ellipse.’ However, many thermodynamic models have notoriously correlated parameters, and this correlation is often quite non-linear. Therefore, we developed the Equal Probability Sampling (EPS) technique that uses the objective function surface itself to sample the parameter space rigorously and directly. The appendix provides a more detailed description of these techniques.

The EPS technique was used by Vasquez et al. [36] to study uncertainty and sensitivity of thermodynamic models. In this work, a case study based on the diisopropyl ether(1) + acetic acid(2) + water(3) liquid–liquid system is presented. Experimental data from Treybal [37] were used to regress the binary interaction parameters for the NRTL equation with $\alpha = 0.3$, using the ASPEN PLUS process simulator. The regression results are presented in Table 3. One hundred samples were generated using both EPS and LHS sampling techniques. To illustrate and study the amplitude of the conditions chosen by the sampling methods, the equilibria based on the extreme values for the composition of component two were plotted for each technique. These results are presented in Fig. 7, which shows a high overestimation of the parameter space produced by the LHS technique, including an unfeasible extreme upper binodal curve. The EPS parameter space estimation is more reasonable from a practical standpoint for both binodal curves.

Table 3

Binary parameters b_{ij} and b_{ji} regressed for the NRTL model for the system diisopropyl ether(1) + acetic acid(2) + water(3)

i	j	b_{ij} (K)	Standard deviation (K)	b_{ji} (K)	Standard deviation (K)
1	2	−400.72	289.84	635.08	240.82
1	3	786.73	34.749	1454.12	50.68
2	3	−343.87	73.68	430.72	381.48

$\alpha = 0.3$.

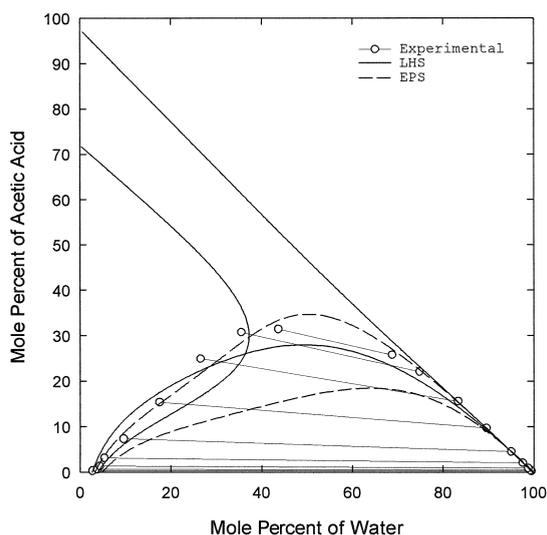


Fig. 7. Extreme equilibrium binodal curves obtained for the sampling techniques LHS and EPS in the LLE prediction of the ternary system diisopropyl ether + acetic acid + water at 25°C. Predicted values by the NRTL model.

A unit operation taken from Treybal [37] was selected to study the sampling effects of EPS and LHS for this problem. The process consists of 8000 kg/h of an acetic acid + water solution, with 30% mass acid, which is countercurrently extracted with diisopropyl ether to reduce the acid concentration in the solvent-free raffinate. The solvent flow rate is 12,500 kg/h, and the column has eight equilibrium stages. The operating conditions are 25°C and 1 atm. The output variable is the percentage of acetic acid extracted at steady state conditions. The operation was simulated using the ChemCAD process simulator.

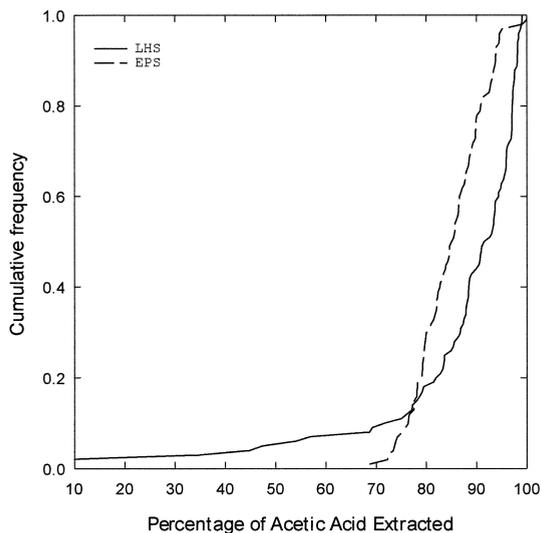


Fig. 8. Uncertainty of percentage of acetic acid extracted using LHS and EPS as sampling techniques in the ternary system diisopropyl ether + acetic acid + water at 25°C, using the NRTL model.

Fig. 8 presents the cumulative frequency curves for the percentage of acetic acid extracted using both sampling techniques. It can be observed that the distribution from EPS is narrower in accordance with the results of Fig. 7. A more realistic or practical interpretation of the uncertainty analysis can be done using EPS than using the LHS approach. The existence of unfeasible simulations is substantially reduced through the EPS technique.

4. Conclusion

Monte Carlo approaches have been shown to be a good tool for sensitivity and uncertainty analyses of thermodynamic models. They can be successfully used to study experimental data source, and data type effects on the regression of binary interaction parameters. Additionally, more information about the regression and the model is generated through the use of the proposed methods. Traditionally, model evaluations are based only on the statistical estimators generated from the regression procedures, which as shown above do not provide a reliable evaluation of the model for practical purposes.

From the results presented, the binary interaction parameters for the NRTL and UNIQUAC models are shown to be very sensitive to experimental data source and data type effects, affecting notoriously the predicted performance of process unit operations. The regression of binary interaction parameters for liquid–liquid applications using binary data (vapor–liquid and liquid–liquid) is more risky than using ternary data (as expected), a conclusion obtained from the uncertainty analysis of predicted process performance of specific processes using the UNIQUAC model.

Experimental uncertainties can lead to significant and unexpected uncertainties in predicted process performance, which also can be revealed through Monte Carlo uncertainty analysis. The cumulative frequency curves can be used for risk avoidance via rational over-design, quantification of the value of obtaining additional data, and reliable evaluation of the uncertainty of the calculated results. Mean or median calculations are insufficient to evaluate the accuracy and precision of simulation and design analyses. Additionally, the users of thermodynamic models tend to over-value the accuracy and precision of these models, making uncertainty analyses even more important.

The proposed EPS technique shows a much more realistic or practical interpretation of the uncertainty present in the regression of thermodynamic models. Resolution problems of techniques like LHS and SHS when dealing with highly nonlinear models are substantially decreased through the use of the EPS approach. The parameter correlation is incorporated directly in EPS rather than using approximated pairing procedures.

More work is required in order to develop methods to quantify the effects of systematic and modeling errors on uncertainty and sensitivity analysis of thermodynamic models. The necessity of improved thermodynamic models and regression techniques to include these issues is demonstrated.

5. List of symbols

b_{ij}, b_{ji}	parameters in the NRTL or UNIQUAC model
\mathbf{H}	Hessian matrix of S
S	maximum-likelihood objective function for regression
T	absolute temperature

\mathbf{V}_θ	covariance matrix of θ
x_{12}	mole fraction of component 2 in phase 1

Greek

α_{ij}	parameter in the NRTL model
θ	vector model parameters
σ	standard deviation

Subscripts

i	index for experimental data points
-----	------------------------------------

Superscripts

e	experimental value
M	model value
T	transpose
*	value at the optimum set of parameters

Acknowledgements

This work was supported, in part, by US National Science Foundation, grant CTS-96-96192.

Appendix A. Equal probability sampling

When dealing with methods to perform sensitivity and uncertainty analysis on thermodynamic models, the sampling approach becomes an important issue for several reasons. Unbiased and accurate samples have to be obtained from the *parameter space* of the model, where the *parameter space* is defined by a *joint probability distribution* between all possible values for the model parameters and the *data space* used to obtain them with known or unknown correlation. Depending on the mathematical behavior of the objective function near the optimum, the *parameter space* can be represented using several approaches. Traditionally, the sampling process is performed using stratified sampling techniques over the individual parameter distributions, and then the correlation is approximated through the use of pairing procedures. The distributions used in these techniques are usually normal linked by the covariance matrix and correlation coefficients. Among the most common stratified sampling techniques available are LHS and SHS. These sampling methods divide the distributions into intervals of equal probability for the case of LHS (see Iman and Shortencarier [32]) or using a determined sequence for the case of SHS (Hammersley sequence points). Then, one sample is taken at random from within each interval for LHS. The sampling points in SHS are defined for the Hammersley sequence (see Kalagnanam and Diwekar [33]). In both methods, the correlation among the parameters is approximated using a variance reduction technique described in detail by Iman and Conover [34]. Mathematically speaking, these approaches are an approximation of the level sets defined by Eq. (2), which is the objective function approximated by an ellipsoid around the optimum using the Taylor's series expansion. The vector of unknown parameters is denoted by θ , and θ^*

represents the optimum parameters. \mathbf{H}^* is the Hessian of $S(\theta)$ at $\theta = \theta^*$, and \mathbf{V}_θ is the covariance matrix of θ .

$$S(\theta) = S^* (\theta - \theta^*)^T \frac{1}{2} \mathbf{H}^* (\theta - \theta^*) \approx S^* + (\theta - \theta^*)^T \frac{1}{2} \mathbf{V}_\theta^{-1} (\theta - \theta^*) \quad (2)$$

When the model involved is highly nonlinear, the right-hand-side of Eq. (2) represents poorly the behavior of the objective near the optimum. We proposed a new resampling method called Equal Probability Sampling (EPS), which automatically adapts to the geometry of the level sets. The level set estimation used is an improved one proposed by Potocký and Ban [35] with a new resampling scheme along the level set. The EPS approach takes into account the nonlinear correlation among the parameters produced by intrinsic curvature and parameter-effects near the optimum of the objective function. This fact produces more realistic samples from the parameter space standpoint (fewer unlikely samples), and a faster convergence of the ‘true’ output distribution. With EPS, the parameter space is bounded using level sets of $S(\theta)$ obtained from higher order terms of the Taylor’s series expansion. The p -dimensional hypersurface (p represents the number of parameters regressed in the model) or level set around θ^* is divided into hypersurfaces of equal probability defined using confidence level values, and one sample is taken at random from the boundary of each hypersurface. Once the level sets are defined, in order to take a random sample from each level set, a uniformly distributed sampling is desired because every value on the boundary of the level set is equally likely for a given confidence level. For the two-dimensional case, the uniform sampling can be done based on arc-length calculations, but for higher dimensions this problem does not have a straightforward solution. Suggestions are made and used based on sampling patterns generated from ellipsoidal approximations of the level sets for the objective function. Vasquez et al. [36] more extensively discuss the algorithms, implementation, and computational issues of the EPS technique.

References

- [1] U.M. Diwekar, E.S. Rubin, *Comput. Chem. Eng.* 15 (1991) 105–114.
- [2] P.D. Chaudhuri, U.M. Diwekar, *Comput. Chem. Eng.* 21 (1997) 733–738.
- [3] K.P. Halemane, I.E. Grossman, *AIChE Journal* 29 (1983) 425–433.
- [4] E.N. Pistikopoulos, M.G. Ierapetritou, *Comput. Chem. Eng.* 19 (1995) 1089–1110.
- [5] M.G. Ierapetritou, E.N. Pistikopoulos, C.A. Floudas, *Comput. Chem. Eng.* 18 (1994) S553–S557.
- [6] M. Streich, H. Kistenmacher, *Hydrocarbon Processing* 59 (1979) 237–241.
- [7] A.R. Nelson, J.H. Olson, S.I. Sandler, *Ind. Chem. Process Des. Dev.* 22 (1983) 547–552.
- [8] S. Macchietto, G. Maduabeuke, R. Szczepanski, *Fluid Phase Equilibria* 29 (1986) 59–67.
- [9] M.E. Reed, W.B. Whiting, *Chem. Eng. Commun.* 124 (1993) 39–48.
- [10] W.B. Whiting, T.-M. Tong, M.E. Reed, *Ind. Eng. Chem. Res.* 32 (1993) 1367–1371.
- [11] W.B. Whiting, *J. Chem. Eng. Data* 41 (1996) 935–941.
- [12] J.M. Sørensen, T. Magnussen, P. Rasmussen, A. Fredenslund, *Fluid Phase Equilibria* 3 (1979) 47–82.
- [13] W.R. Esposito, C.A. Floudas, *Ind. Eng. Chem. Res.* 37 (1998) 1841–1858.
- [14] R.C. Reid, J.M. Prausnitz, B.E. Poling, *The Properties of Liquids and Gases*, 4th edn., McGraw-Hill, New York, 1987.
- [15] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, New York, 1987.
- [16] C.F. Dietrich, *Uncertainty, Calibration and Probability*, 2nd edn., Adam Hilger, Bristol, 1991.
- [17] J.P. Novák, J. Matouš, J. Pick, *Liquid–Liquid Equilibria*, Elsevier, New York, 1987.
- [18] V.R. Vasquez, W.B. Whiting, *Fluid Phase Equilibria* 142 (1998) 115–130.

- [19] J.M. Sørensen, W. Arlt, Liquid–Liquid Equilibrium Data Collection, DECHEMA, Frankfurt am Main, 1980.
- [20] A.V. Brancker, T.G. Hunter, A.W. Nash, *J. Phys. Chem.* 44 (1940) 683–698.
- [21] W.D. Bancroft, S.D. Hubbard, *J. Am. Chem. Soc.* 64 (1942) 347–353.
- [22] W. Reinders, C.H. De Minjer, *Recl. Trav. Chim. Pays-Bas* 66 (1947) 573–604.
- [23] B.F. Ruiz, R.D. Prats, *Fluid Phase Equilibria* 10 (1983) 77–93.
- [24] B.D. Smith, *Design of Equilibrium Stage Processes*, McGraw-Hill, New York, 1963.
- [25] D.S. Abrams, J.M. Prausnitz, *AIChE Journal* 21 (1975) 116–128.
- [26] Aspen Technology, *ASPEN PLUS Reference Manual*, Vol. 2, Cambridge, MA, 1996.
- [27] V.R. Vasquez, W.B. Whiting, *Ind. Eng. Chem. Res.* 37 (1998) 1122–1129.
- [28] L.V. Kogan, I.V. Deizenrot, *Zh. Prikl. Khim. (Leningrad)* 48 (1975) 2757–2759.
- [29] A. Apelblat, A. Tamir, M. Wagner, *Fluid Phase Equilibria* 4 (1980) 229–255.
- [30] H.E. Eduljee, V.N. Kumarkrishnarao, M.N. Rao, *J. Chem. Eng. Data* 3 (1958) 44–50.
- [31] R.M. Stephenson, *J. Chem. Eng. Data* 37 (1992) 80–95.
- [32] R.L. Iman, M.J. Shortencarier, *A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use with Computer Models*, Report NUREG/CR-3624, National Technical Service, Springfield, VA, 1984.
- [33] J. Kalagnanam, U.M. Diwekar, *AIChE Journal* 43 (1997) 440–447.
- [34] R.L. Iman, W.J. Conover, *Commun. Stat.-Simula. Comput.* 11 (1982) 311–334.
- [35] R. Potocký, R.V. Ban, *Applications of Mathematics* 37 (1992) 29–39.
- [36] V.R. Vasquez, M. Meerschaert, W.B. Whiting, submitted for publication to *Risk Analysis*, 1999.
- [37] R.E. Treybal, *Mass-Transfer Operations*, 3rd edn., McGraw-Hill, New York, 1981.