

LARGE-SCALE DATA REGRESSION FOR PROCESS CALCULATIONS

Agarwal, R. K., Li, Y.-K., Santollani, O. J. and Satyro, M.A. *

Virtual Materials Group, 657 Hawkside Mews NW, Calgary, AB, Canada T3G 3S1

Paper to be presented at the "International Conference on Distillation and Absorption", September 20 – October 2, 2002, Kongresshaus Baden-Baden, Germany

Abstract

Accurate vapor and liquid fugacities are fundamental quantities for the calculation of reliable vapor-liquid equilibrium and consequently they have a direct impact on sizing of separation equipment as well as on the quality of simulations used to understand plant behavior. Traditionally, vapor pressures are taken from a given data source and many times engineers take interaction parameters for equations of state or activity coefficient models from the literature. This approach is prone to error and inconsistency, sometimes creating unrealistic simulations showing non-existent azeotropes, incorrect phase behavior or even creating azeotropes between isomers. Perhaps more important, the parameters published in the literature do not include uncertainty information, and quality assurance on the obtained results is not possible.

Since 1999 Thermodynamics Research Center (TRC) and Virtual Materials Group (VMG) are working together towards the creation of state of the art databases of model parameters for the calculation of physical properties necessary for the modeling of complex chemical processes. Our strategy follows a step by step development path where extensively validated software building blocks are produced and assembled to be used in multiphase flash calculations. This strategy was built upon the following principles:

- 1 The vapor pressures for pure components are derived from a single data compilation and the vapor pressure parameters such as Antoine constants or cubic equation of state attractive term parameters are derived using a single set of recommended vapor pressure data taking into account recommended uncertainties for each vapor pressure data point.

- 2 Pure component parameters necessary for phase modeling such as critical properties and liquid molar volumes are collected from evaluated values stored in the TRC SOURCE data system together with estimated uncertainties.

3 Pure component parameters and extensive VLE data are used to generate a large collection of interaction parameters for each available dataset. Each dataset is regressed using the most rigorous objective function available depending on the dataset type. Special care for modeling azeotropic points is taken and all the calculations are performed using rigorous multiphase flashes. Complete error analysis for each dataset is performed, and detailed validations are generated for every dataset. Separate interaction parameter collections are calculated for each vapor phase model / activity coefficient model combination. In addition, collections of interaction parameters for GE (Excess Gibbs Energy) - based equations of state and quadratic mixing rule- based equations of state are generated.

4 After all the available VLE data are regressed, specially developed ranking system automatically examines each binary pair ensemble of datasets and the most adequate set for distillation calculations is selected. Subsequently, internally structured databases of interaction parameters with error information and bibliography are created.

The final product of this data regression effort is a consistent set of interaction parameters, vapor pressure and pure component parameters that can be used confidently for process calculations, based on extensive experimental data as well as extensive data quality evaluation, combined with full backtracking of data origin and model fitness. The benefits of this fundamental approach to data regression are internal consistency, quality assurance, extensive flash engine test and ability of ranking and suggesting thermodynamic models based on the makeup of mixtures of interest.

Introduction

The most fundamental aspect of process simulation calculations can be traced back to the related thermodynamic model used for the mixture phase behavior and physical properties. Underneath this thermodynamic model we invariably find pure component properties such as critical properties and vapor pressures, as well as interaction parameters necessary to guarantee adequate performance of mixture models such as UNIQUAC and Peng-Robinson.

Common engineering practice stresses the convenience of computer calculations. However, more often than not the accuracy of the underlining thermodynamic method used for simulations is not accessed, or, at best, accessed in an incomplete manner since most of the data collections currently available do not provide complete statistical information necessary for the proper performance evaluation of a model. For example, good quality engineering databases such as DIPPR 801 [1] and Yaws [2] provide only fragmentary statistical information about correlation errors in the form of average and maximum errors for temperature dependent properties such as vapor pressure. Standard deviations for important temperature independent physical properties such as critical properties and acentric factors are not provided. Temperature dependent properties correlations are also incomplete in a sense that the variances for regressed parameters are not available, thus precluding the use of such data for a comprehensive error propagation analysis of calculated physical properties for mixtures.

For mixture interaction parameters, the situation is even more confusing. Interaction parameters for VLE models are indirect functions of pure component properties, and have to be uniquely identified with these pure component parameters. In the databases currently used to support VLE calculations there is no explicit guarantee that the vapor pressure and other pure component parameters are consistent with the interaction parameters being used in the simulations [3, 4, 5, 6, 7, 8] with the exception of one property data package system [9] that assures automatic trace back capabilities to all of its interaction parameters and pure component properties. This lack of quality control and traceability creates the potential for inconsistent simulations that “look” reasonable, but may have basic errors that are not obvious, and may cause significant economic losses related to the equipment performance, product specification failure and plant operability problems. These are known issues in chemical process simulation, but since their proper resolution depend on significant resources related to data management, expert analysis, and processing capabilities, so far, these problems have never been resolved in a comprehensive manner.

Our goal is to create a consistent data source for chemical process simulation using the most accurate and well maintained data source available, combined with extensive quality control, consistent use of thermodynamic quality tools such as thermodynamic consistency tests, extensive use of process simulation tools such as multiphase flash calculations, and complete statistical data analysis for the final estimation of uncertainties of the calculated values. It is our belief that a numerical results provided by a process simulation software tools without quantifiers of their uncertainties are incomplete from a scientific point of view, and potentially dangerous when used by engineers without appreciation for error estimates in the underlining thermodynamic models as well as errors in the underlining property data used to power the thermodynamic model such as vapor pressures and interaction parameters.

This quality-driven solution for the calculation of physical properties used in process simulation is possible only if a robust data management system of pure component and mixture data with evaluated uncertainties is available [10], and the necessary thermodynamic framework of models and multiphase flash calculations is developed [9] to be used for extensive data regression and statistical analysis within the structure of rigorous flash calculations used for process simulation.

Theoretical Framework

The theoretical framework necessary to achieve successfully the goal of quality-driven process simulation can be divided into three broad categories as database computer technology, statistical framework and thermodynamic modeling technology.

Database and Computer Technology

The choice of a proper database to work, as the backbone of the data regression and statistical analysis is the most fundamental aspect of this project and it was the object of extensive study. We have chosen to use a combination of several electronic databases developed by the Thermodynamics Research Center at the National Institute of Standards and Technology in Boulder, Colorado as the source of pure

component and mixture information for the generation of the necessary statistical information for uncertainty estimates in process calculations. The processing of information from these databases, in turn, provides the information stored in our own physical property database. The TRC databases are summarized in the list below [10 – 15].

- NIST Standard Reference Database 85-NIST/TRC Table Database WinTable, Version 2002 (TABLES)
- NIST Standard Reference Database 87-NIST/TRC Vapor Pressure Database Version 2001 (VP)
- NIST Standard Reference Database 88-NIST/TRC Compounds Database Wincmp, Version 2002 (CMP)
- NIST Standard Reference Database 100-NIST/TRC Density Database TrcDensity, Version 2002 (LD)
- NIST/TRC Standard Reference Database 91-Floppy Book on Viscosity of Liquids (LVIS)
- NIST/TRC Standard Reference Database 92-Floppy Book on International Data Series Selected Data on Mixtures 1973-2001 (IDS)
- NIST/TRC Standard Reference Database 93-NIST/TRC Floppy Book on Vapor-Liquid Equilibrium Data for Binary and Ternary Systems (VLE)
- NIST/TRC Standard Reference Database 96-NIST/TRC Floppy Book on Heats of Mixing (HE)

These databases are the kernel for the current work being done at Virtual Materials Group for the creation of a standard database of pure component and mixture data for process simulation, named the VIRTUAL database. Some of the databases (such as the TABLE database above) do not contain uncertainty information explicitly expressed, but rather contain recommended values of physical properties based on evaluated data where the number of the significant digits can be used as a measure of the uncertainties. TRC's largest data system of 'raw' experimental data with evaluated uncertainties, the SOURCE data system, contains the majority of the thermophysical data published [11, 15].

It is important to note that the work involved in the creation of VIRTUAL is not static, but rather continuous based on new revisions of the recommended data with the inclusion of recently reported information. This need for continued revisions to ensure that the most accurate data and statistical information is provided for process calculations involve a significant amount of software engineering to ensure that changes in the database are properly logged and that users can understand why simulation results may have changed with the use of calculations based on different versions of VIRTUAL. Changes from version to version are also catalogued in an automatic report, thus allowing users to critically review their designs and invest engineering time revising solutions if necessary.

The interaction of all the databases is described in figure 1.

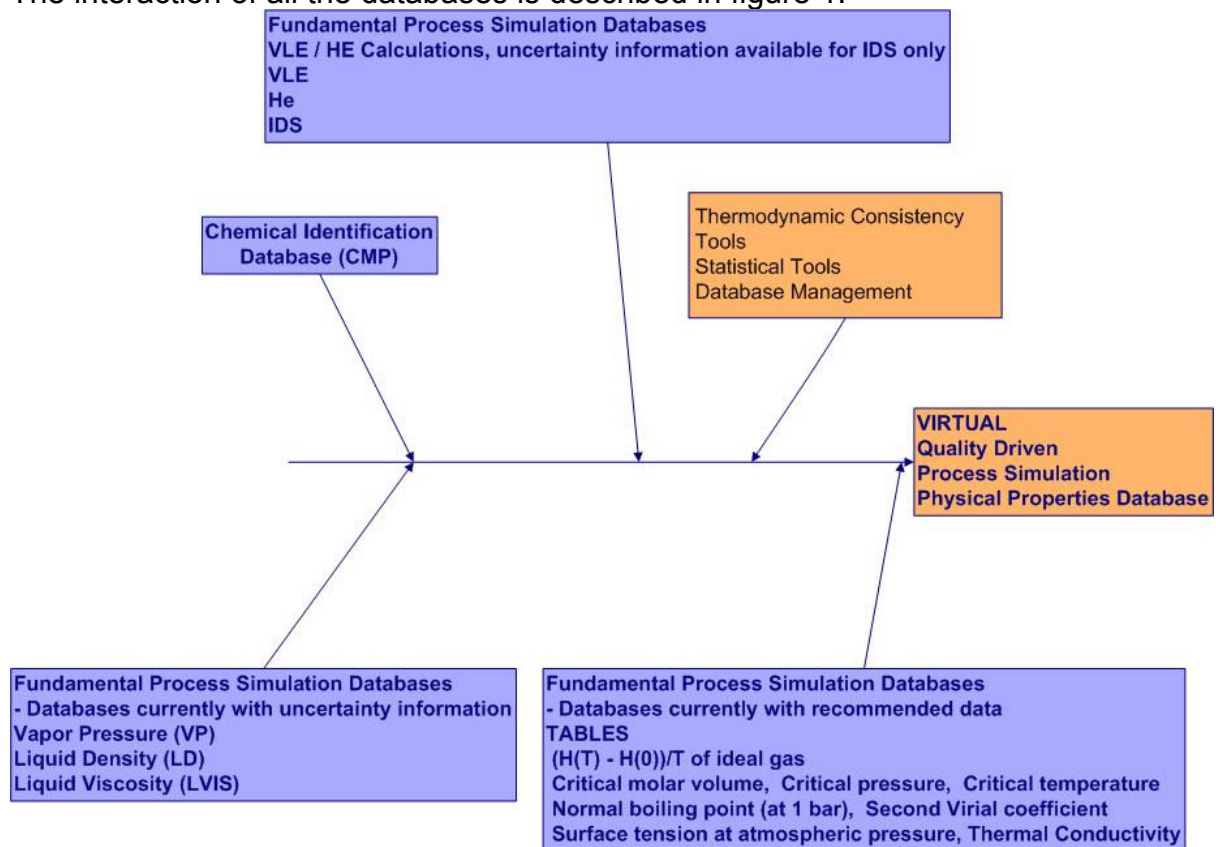


Figure 1. VIRTUAL database integration structure

All the databases used in the creation of VIRTUAL as well as VIRTUAL itself are in relational format. VIRTUAL supports several formats for data extraction independent of its own native relational table using software interfaces. These interfaces allow VIRTUAL to expose data for process simulation tools such as Virtual Materials and consequently process simulators such as HYSYS and PRO/II. In addition, users familiar with physical property databases such as DIPPR 801 can explore VIRTUAL data using VIRTUAL's software interfaces. For users who have their own data formats a data dictionary is provided for the creation of custom interfaces without the need of programming (figure 2).

Currently all databases are maintained using Microsoft Access and the main data regression and data maintenance machine is a Pentium 4 computer operating at 1.8 GHz with 80 GB disk space.

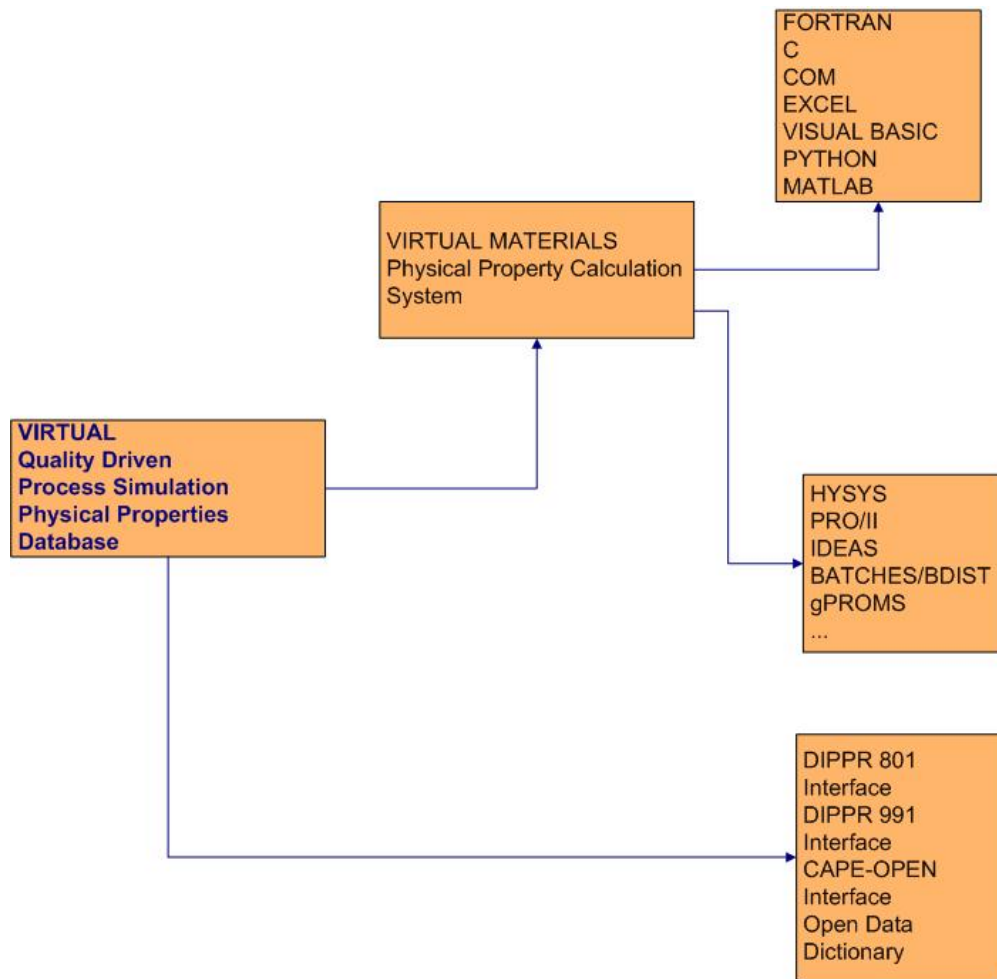


Figure 2. Interfaces for data access from VIRTUAL database. Note that specific users can access data either directly or through the process simulation tools.

The simple collection of data is certainly not enough to accomplish a truly useful quality-driven database for process simulation. Careful data selection and back tracking capabilities are as fundamental as the data themselves to allow users designing actual process equipment to check the quality of results. This will be illustrated in the Thermodynamic Modeling Technology section.

Statistical Tools

The necessary statistical tools are defined by the specifications we imposed onto the VIRTUAL database for performing the services related to providing the requested information for the evaluation of uncertainties during process simulation calculations. It is important to briefly review what is involved in the fulfillment of this goal from a mathematical point of view.

If one assumes that the statistical error on a value is small and we are measuring a function of several variables where \vec{x} represents the vector of random variables whose true values are given by the vector \vec{X} , we can represent a true value of a derived quantity by equation 1:

$$U = f(\vec{X}) \quad (1)$$

One is naturally faced with the need of defining how small is small. Mandell [16] suggests that if the errors are in the order of 10% or smaller of the variable of interest then the law of propagation of errors can be reliably used.

Now let us define the vector $\vec{\varepsilon}$ as the vector containing statistically independent errors of \vec{x} . Then the error in U , denoted as δ and resulting from the errors $\vec{\varepsilon}$ has a variance described by equation 2.

$$V(\delta) = \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i} \right)^2 V(\varepsilon_i) \quad (2)$$

Therefore, if we want to be able to estimate the error in the calculation of a thermodynamic property, we need to know not only the values of the parameters it depends upon, but also their estimated variances. For example, if one is calculating the composition of a component using Raoult's law, then it is necessary to know the uncertainty of the parameters used in the vapor pressure correlation. In this simple scenario, we can assume that the thermodynamic behavior of the system is well described by an ideal solution model, and therefore the vapor composition can be described by the equation 3 (y_i is a mole fraction of the component i in a vapor phase, x_i is a mole fraction of the component i in a liquid phase; Pv_i is a partial vapor pressure of the component i , P is a total pressure; A and B are parameters defined in the equation 5, T is a temperature).

$$y_i = \frac{Pv_i}{P} x_i = f(Pv_i, P, x_i) = g(A_i, B_i, T, P, x_i) \quad (3)$$

Then, the uncertainty in the composition of component i in the vapor is given by equation 4.

$$V(\delta_{y_i}) = \left(\frac{\partial g}{\partial A_i} \right)^2 V(\varepsilon_{A_i}) + \left(\frac{\partial g}{\partial B_i} \right)^2 V(\varepsilon_{B_i}) + \left(\frac{\partial g}{\partial T} \right)^2 V(\varepsilon_T) + \left(\frac{\partial g}{\partial P} \right)^2 V(\varepsilon_P) + \left(\frac{\partial g}{\partial x_i} \right)^2 V(x_i) \quad (4)$$

The vapor pressure of component i can be represented by the integrated Clausius-Clapeyron equation (equation 5).

$$Pv_i = \exp\left(A_i + \frac{B_i}{T}\right) \quad (5)$$

Equation 4 clearly shows why it is important to collect complete statistical information about all parameters used for the calculation of a physical property used for process simulation. The uncertainties of the parameters A_i and B_i , in turn, could only be determined consistently on the basis of the uncertainties of the experimental values of vapor pressure of the component i used to calculate A_i and B_i .

All software subroutines developed for the data regression are written using Microsoft's Visual Basic language as ActiveX components. The mathematical techniques for the determination of model parameters are well established [17-19].

Basically, the non-linear regression engine uses a modification of Marquart-Levenberg algorithm combined with singular-value decomposition [19]. In addition to parameter uncertainties the software stores chi-square function values, significance test results, average and maximum absolute percent errors, average and maximum absolute deviations and bias, as well as information about thermodynamic calculation success or failure for each data point used in the calculations. Naturally, all the thermodynamic calculations are performed through a multiphase flash engine thus ensuring that not only the data but the flash engine be extensively tested. For example, during the determination of optimal VLE parameters for an activity coefficient model, more than 20,000 binary datasets are evaluated and more than 300,000 flash calculations are performed.

Thermodynamic Modeling Technology

Usually the development of physical property databases for process simulation is performed completely independent of a process simulation engine development. In our judgment, that is a significant drawback. Coherent development of the property database and simulation engine provides an opportunity to extensively test the algorithms under a vast space of test cases while determining optimal parameters for activity coefficient models as well as equations of state. In turn, sometimes discrepancies between model and data uncovers problems with the data, such as typographical errors, unit conversion errors, ordering errors (in the case of binary or multicomponent data) and incorrect literature values.

Our thermodynamic modeling is based on Virtual Materials Group Virtual Materials™ property package framework. The Virtual Materials™ framework contains all commonly encountered thermodynamic models used for process simulation in addition to sophisticated models used for the simulation of reactive systems, coupled with a rigorous multiphase flash and software interfaces that allow it to be run in line with a variety of existing application software. These software interfaces make it easy to access model parameters inside the property package and are used extensively when determining the physical property parameters in the VIRTUAL database.

The ability to run the thermodynamic engine without a specific graphical interface but rather by just accessing its parameters and starting calculations is extensively used in the creation of the VIRTUAL database where Virtual Materials™ is called many times during the optimization of model parameters. This, in turn, is also useful for the flash algorithm testing since during the early stages of optimization, model parameters vary significantly and do not correspond to physically meaningful values.

Prototyping – VLE Interaction Parameter Regression

As a prototype for the VIRTUAL database creation, the determination of interaction parameters for all activity coefficient models supported by Virtual Materials™ was done in 2000 and 2001. This extensive data regression and data selection served as

a foundation for a creation of the VIRTUAL database and provided a preliminary high quality database of interaction parameters for activity coefficient models.

The generation of the default interaction parameter databases was performed according to the following procedure:

- 1 Select thermodynamic model (e.g. NRTL / Ideal Gas)
- 2 Determine optimal interaction parameters by minimizing the bubble pressure objective function for all datasets present in the VLE database (NIST Standard Reference Database 93) using Virtual Materials rigorous multiphase flash system.
- 3 Determine thermodynamic consistency and other quality parameters related to each dataset.
- 4 Based on the quality of the regressions for each dataset for a specific binary mixture and the quality parameters of each dataset related to that binary mixture choose a default dataset.
- 5 Store default binary parameters in VIRTUAL database.
- 6 Calculate bubble and dew pressure or temperature curves as a function of a composition using default binary interaction parameters to verify accuracy of parameters and reliability of flash engine.
- 7 Repeat a procedure for all thermodynamic models supported by Virtual Materials Group.

The bubble pressure used in the optimization procedure is defined by equation 6.

$$O^2 = \frac{1}{\sigma_p^2(N-n)} \sum_{i=1}^{np} (P_i^e - P_i^c)^2 \quad (6)$$

Here, σ_p is the standard deviation in pressure, assumed to be equal to 0.133 kPa, P^e is the experimental bubble pressure, P^c is the calculated bubble pressure, N is the total number of points for the dataset and n is the number of optimized parameters used by the model.

The bubble pressure is calculated using the following expression for the vapor and liquid fugacities, equation 7 (ϕ_i , γ_i , and ϕ_i^s are a fugacity coefficient, an activity coefficient, and a fugacity coefficient at saturation pressure for the component i).

$$\phi_i y_i P = \gamma_i x_i \phi_i^s P_o y_i P_{vi} \quad (7)$$

The fugacity coefficients are calculated using any of the equations of state provided in the Virtual MaterialsTM framework, while the activity coefficients are calculated using any of the activity coefficient models supported by Virtual MaterialsTM. These are summarized in the tables below.

Table 1. Virtual Materials™ Liquid Phase Models

Activity Coefficient Models
Ideal Liquid (not used in regressions)
Regular Solution (not used in regressions)
Margules
Margules 4 Suffix
Monsanto van Laar
Wilson
Wilson-Tsuboka-Katayama
NRTL
UNIQUAC
UNIQUAC-Sander

Table 2. Virtual Materials™ Vapor Phase Models

Fugacity Coefficient Models
Ideal Gas
Virial Equation of State (Hayden-O'Connell)
Redlich-Kwong
Soave-Redlich-Kwong
Refinery-SRK (SRK using API's definitions)
ZSRK (SRK optimized for natural gas compressibility factor calculations)
Peng-Robinson
Advanced Peng-Robinson
NGL-Peng-Robinson (PR using GPA's RR-28 specifications)
GE-Peng-Robinson (PR using Huron-Vidal mixing rules and Mathias-Copeman attractive term for pure components)
PSRK – Predictive SRK

The complete interaction parameter database contains interaction parameter tables for 96 different combinations of vapor and liquid phase models, or approximately 288,000 recommended interaction parameters for approximately 3,000 different binary pairs.

The fact above has the important implication of forcing consistent calculations for all vapor models used for activity coefficient-based models. For example, the DECHEMA VLE data collection [21] has an extensive collection of interaction parameters for a few activity coefficient models assuming ideal gas behavior. The VIRTUAL database contains 9 times more information since for each vapor model used in conjunction with an activity coefficient model, a different database of interaction parameters and recommended interaction parameters is generated. For example, when determining optimal interaction parameter values for the NRTL activity coefficient model, the following interaction parameter databases are generated:

Table 3. Different NRTL-type databases generated for each vapor phase model

Database	Vapor Model	Liquid Model
NRTL-1	Ideal Gas	NRTL
NRTL-2	Virial (Hayden-O'Connell)	NRTL
NRTL-3	Redlich-Kwong	NRTL
NRTL-4	Soave-Redlich-Kwong	NRTL
NRTL-5	Refinery-SRK	NRTL
NRTL-6	ZSRK	NRTL
NRTL-7	Peng-Robinson	NRTL
NRTL-8	Advanced Peng-Robinson	NRTL
NRTL-9	NGL-Peng-Robinson	NRTL
NRTL-10	GE-Peng-Robinson	NRTL
NRTL-11	PSRK	NRTL

This procedure ensures consistency between the interaction parameters used for thermodynamic calculations and pure component physical properties. Also, since the standard Virtual Materials multiphase flash engine is used for all calculations, the results are consistent with what one would expect when using the interaction parameters for process calculations.

Since the number of equations of state in Virtual Materials is continuously increasing, the number of interaction parameters for activity coefficient models is correspondingly upgraded.

Thermodynamic Consistency and Other Quality Parameters

The thermodynamic consistency of a dataset is determined based on the point consistency test as suggested by Gess [20]. For each binary, optimal interaction parameters for the four-suffix Margules equation are determined by minimizing the bubble pressure objective function (equation 6) as it follows from equations 8 and 9 (A, B, and D are parameters of the four-suffix Margules equation).

$$\ln \gamma_1 = x_2^2 [A + 2(B - A - D) - 3Dx_1^2] \quad (8)$$

$$\ln \gamma_2 = x_1^2 [B + 2(A - B - D) + 3Dx_2^2] \quad (9)$$

Using the optimized interaction parameters for the Margules 4-suffix equation the vapor compositions can be computed as:

$$y_i = \frac{P_{vi} \gamma_i x_i}{P} \quad (10)$$

Then, the average error and bias in vapor composition are calculated:

$$\Delta y = \frac{1}{N} \sum_{i=1}^{np} |y_1^c - y_1^e| \quad (11)$$

$$Bias = \frac{1}{N} \sum_{i=1}^{np} y_1^c - y_1^e \quad (12)$$

The dataset is considered consistent if the average error in vapor composition is less than 0.01.

The bias can be used to check for systematic deviations in the calculated vapor composition.

VIRTUAL also includes the estimated classification of the binary systems based on the magnitude of the Margules two-suffix equation parameters. This provides a magnitude of the maximum Gibbs free energy of the binary mixture, and it is useful as a 'rough' indication of the non-ideality of a system. Following the suggestion by Gess and coworkers [20], we classify systems with a two-suffix interaction parameter equal or less than 0.6 as nearly ideal. Systems having this interaction parameter greater than 0.6 are classified as non-ideal. In addition, the standard deviation with respect to the Margules two-suffix model is calculated, thus providing an estimate of the asymmetry of the Gibbs free energy curve with respect to a mole fraction of the component in the mixture:

$$\sigma_{m2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{np} \left(\frac{g_c^E}{RT} - \frac{g_e^E}{RT} \right)^2} \quad (13)$$

where g^E is the Gibbs free excess energy of the mixture.

Finally, the composition momentum is calculated. The composition momentum describes how well spread the data are with respect to their distribution along the liquid mole fraction axis. Data sets with data distributed over the liquid compositions for a binary mixture have a composition momentum nearly equal to zero.

Regression Statistics

The following information is collected during the data regression for each dataset.

For all dataset types

- a) Absolute average error in temperature and in pressure;
- b) Absolute deviation in pressure and temperature;
- c) Maximum error in temperature and pressure;
- d) Maximum deviation in temperature and pressure.

For TXY or PXY or TPXY dataset types

- a) Absolute average error in vapor composition;
- b) Absolute deviation in vapor composition;
- c) Maximum error in vapor mole fraction;
- d) Maximum deviation in vapor mole fraction.

Average errors are defined as:

$$a = \frac{100}{N} \sum_{i=1}^{np} \left| \frac{m_i^c - m_i^e}{m_i^e} \right| \quad (14)$$

Average deviations are defined as:

$$a = \frac{1}{N} \sum_{i=1}^{np} |m_i^c - m_i^e| \quad (15)$$

Default Interaction Parameter Selection Procedure

The selection of default interaction parameters for thermodynamic models is almost always biased towards a particular class of applications, the available data and the experience of the user of the data. When creating default interaction parameter databases, we strive to have the defaults selected using uniform criteria, thus allowing to compare the performance of different thermodynamic models and choose the most adequate model for particular application. A software component designed to work under any host program was developed to allow for identification of potential problems with the default parameters and providing information necessary for making decisions about collecting new experimental data and regressing new interaction parameters that may be more adequate for a specific application.

Selection Process

After all the VLE data is processed for a given thermodynamic model, each data set is ranked. For VLE data at or close to normal atmospheric pressure, the following empirical criteria are used:

- 1 Only data sets with corresponding function (equation 6) values less than 1000 are considered fit for selection. This corresponds to an average number of points per dataset equal to 10, an error per point of approximately 1 kPa and an error in pressure of about 0.1 kPa. These values are certainly empirical, but they seem to represent most of the cases of interest (except for systems at high vacuum)
- 2 For sets that pass criteria 1, thermodynamic consistency is checked. If the data are consistent, the ranking function is set to zero, else it is set to 1. If the dataset is of type *TXY*, zero is added to the ranking function, if the type is *TX*, 1 is added to the ranking function, if the type is *PXY*, 2 is added to the ranking function, otherwise 3 is added to the ranking function.
- 3 The average of the absolute difference between the pressures for each point in the dataset and 101.325 kPa is added to the ranking function.
- 4 The average deviation in temperature calculated using the selected thermodynamic model is added to the ranking function.
- 5 The absolute value of the composition momentum is added to the ranking function.
- 6 The dataset with smallest ranking function value is selected as the default dataset for the binary pair, and the corresponding interaction parameters are set as default.

Selection Process Characteristics

The selection process was designed to choose data sets of quality from a thermodynamic point of view, with good overall model fitting, close to atmospheric pressure and with data reasonably spread over the composition space. These are generally good principles for defaulting data sets and corresponding binary coefficients for distillation applications at modest pressures. Naturally, different processes may require a reevaluation of the parameters, for example, if a superfractionator model is desired it may be more adequate to change the objective function to force a better relative volatility fit instead of a better bubble pressure fit.

Nevertheless, the fact remains that the user of the model can determine at any time the source and the assumptions used for the determination of the physical property parameters for the process calculations. Typical results are shown in figures 3 and 4.

It is interesting to note that using the database structure and thermodynamic tools described, a complete interaction parameter determination involving 24,000 binary pairs based on a bubble pressure objective function and rigorous multiphase flash calculations, and the corresponding selection of a recommended dataset for each available binary pair can be done in approximately 8 hours.

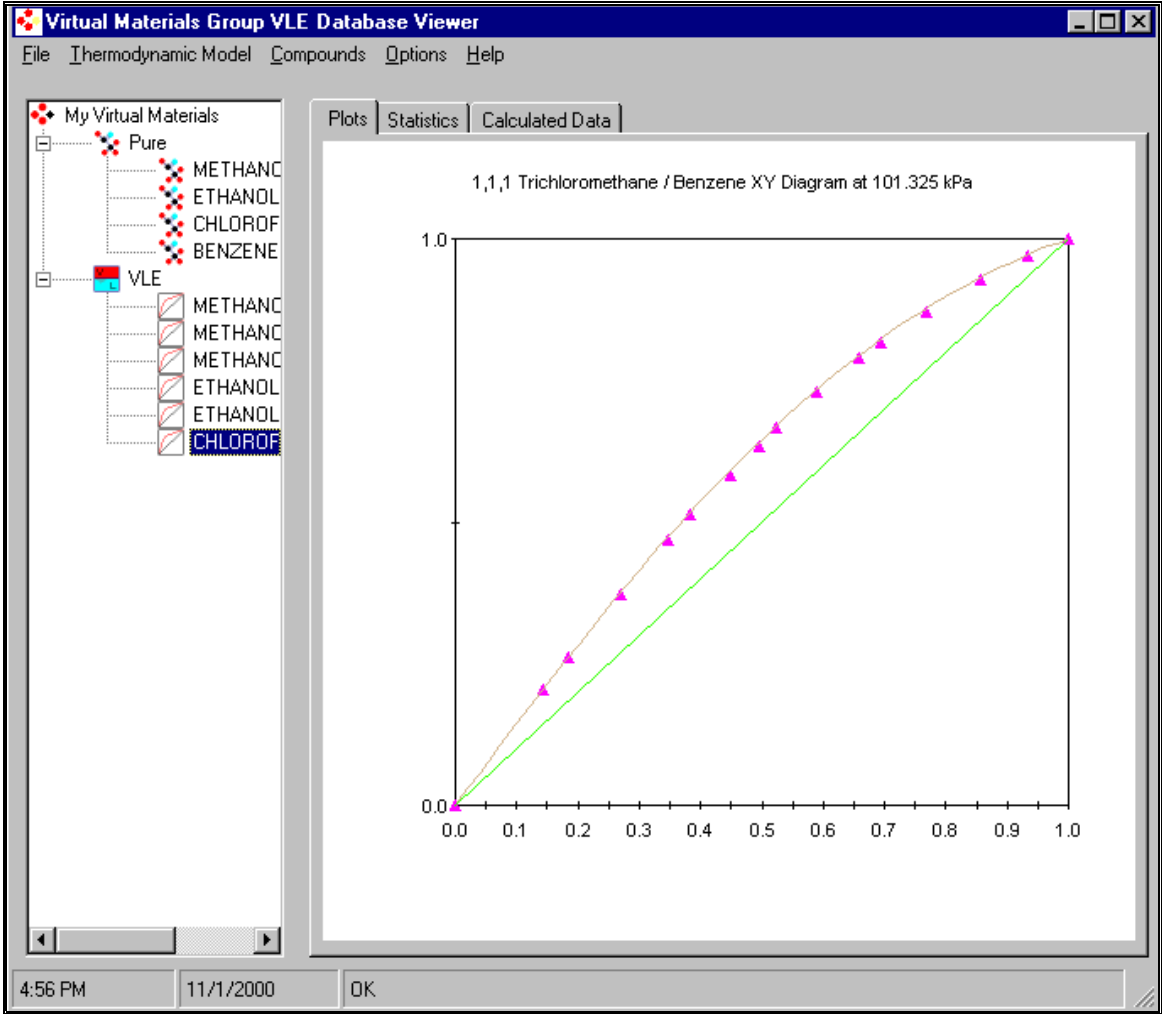


Figure 3. VIRTUAL regression results for 1,1,1-Trichloromethane / Benzene at 101.325 kPa

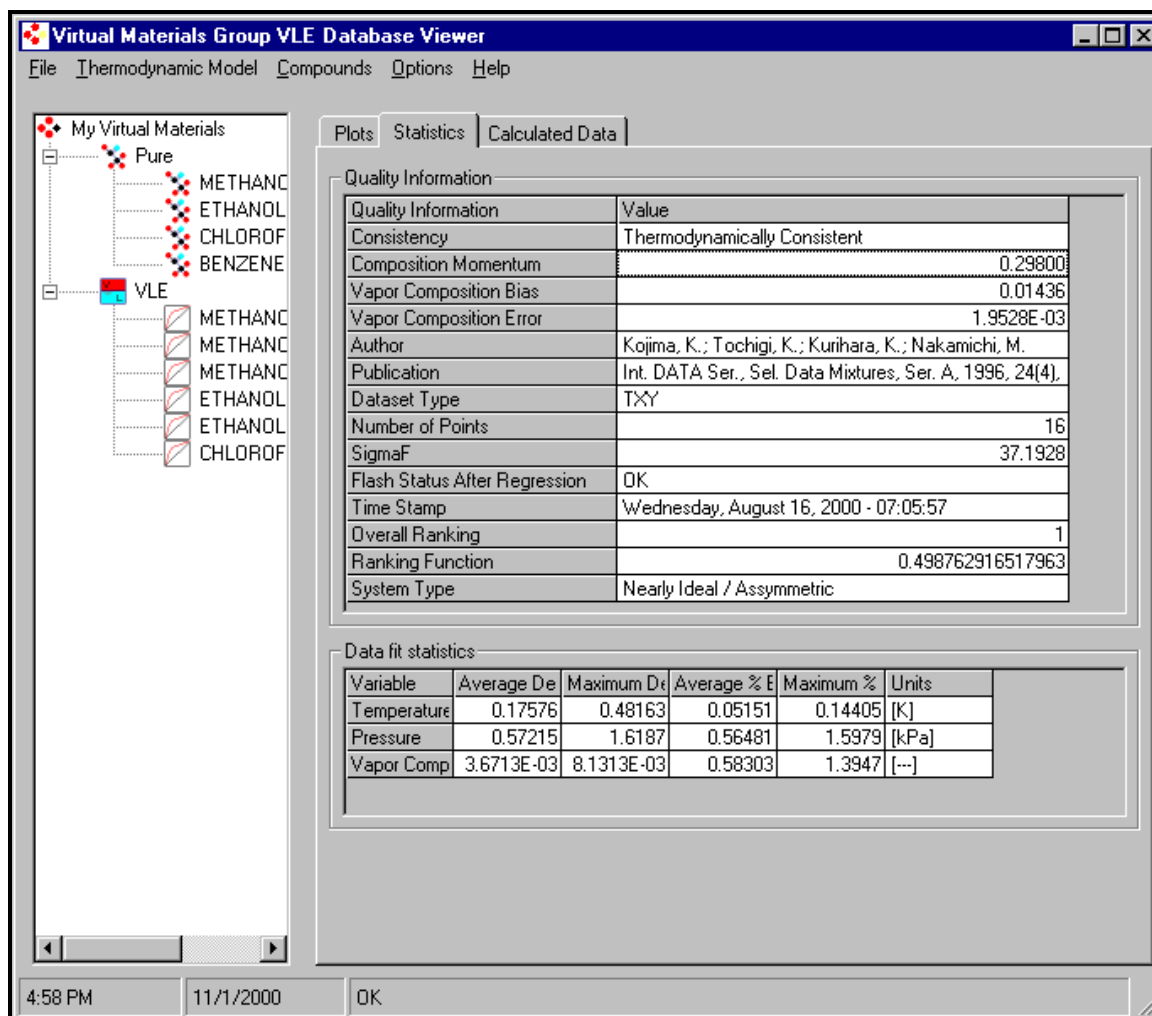


Figure 4. VIRTUAL regression results for 1,1,1 Trichloromethane / Benzene at 101.325 kPa – statistical details

Discussion

The creation of a prototype VIRTUAL database using vapor-liquid equilibrium data was successful, and served to completely test the thermodynamic models used for calculations, flash engine, statistical calculation procedures and database maintenance tools.

With these fundamental components tested, the work on the creation of a definitive VIRTUAL database is in progress where all physical property parameters important for process simulation calculations are being reworked with the necessary care for the collection of statistical information on parameter uncertainty to allow property uncertainty calculations on demand.

The VIRTUAL definitive database currently contains all the pure component data present in the NIST/TRC TABLE, VP, LDN and LVIS databases, and the VLE tables are being reworked to include excess enthalpy data as well as estimated uncertainties in the interaction parameters. This definitive database of thermodynamic data for process calculations will be available in the summer of 2002.

Conclusions

The ability of qualifying the results coming from process simulation calculations is a fundamental step towards a sound basis for the design of chemical equipment and simulation of chemical processes under quantifiable, numerical parameters. Currently this ability is not present in process simulation software and therefore when designing equipment the engineer is faced with the need of adding safety factors based on experiment. The actual value of such safety factors is not based on sound science but rather, it is based on empirical knowledge gathered from comparing designed and actual performance of chemical equipment.

It is natural that a certain degree of art will always be related to chemical equipment design, and our work removes only one factor related to calculation uncertainties, albeit a very significant one since thermodynamic calculations are the basis for most process designs. It is our hope that this same framework can be put to work towards quantifying uncertainties in mass transfer calculations, heat transfer coefficients and other relationships used by engineers, and, in turn, help put the safety factors used for design under a more scientific basis.

Perhaps as important, it is common now to design processes under very tight concentration specifications due to environmental concerns. The knowledge of at least estimated errors in computed thermodynamic variables would be priceless for the proper assessment of effectiveness of processes and for helping to analyze them under realistic parameters. Also, the knowledge of uncertainties can provide actual information about realistic values to be used for process control to avoid waste of time trying to optimize or control process variables that can not be calculated with the accuracy one would associate with results from calculations where uncertainties are not available as normally done today when optimizations are done from a purely mathematical perspective, neglecting the fact that the search values within an error band are for all practical purposes identical.

The VIRTUAL database combines the best available physical property data with extensive evaluation and testing under actual process simulation conditions and is a significant step towards quality-driven process simulation.

Bibliographic References

1. Oscarson, J. L.; Rowley, R. L.; Wilding, W.V. Design Institute for Physical Properties Project 801; AIChE, 2002.
2. Yaws, C.L. Chemical Properties Handbook, McGraw-Hill, New York, 1999.
3. Agarwal, R.; Li Y.-K.; Santollani, O. J.; Satyro, M. A.; Vieler, A. Chem. Eng. Progress, 97(5), 42, 2001; 97(6), 64, 2001.
4. Yang, N. S.; Xu, Z. P.; Chuang, K. T. Hydrocarbon Processing, (5), 109, 1999.
5. Sadeq, J.; Duarte, H.A. and Serth, R.W.; Chemical Engineering Education, (4), 47, 1997.
6. Benso, V.; Bertucco, A. RICHMAC Magazine, Marzo/Aprile, 46, 1995

7. Moura, C. A. D.; Carneiro, H. P. B. Tech. Petrobras, Rio de Janeiro, 34(3/4), Jul/Dez., 17, 1991
8. Zudkevitch, D.; Phase Equilibria and Fluid Property in the Chemical Industry Conference, Berlin, 1980.
9. Virtual Materials Property Package System Users Manual, Virtual Materials Group, Inc., Calgary, Alberta, Canada, 2001.
10. Yan, X.; Dong, Q.; Frenkel, M.; Hall, K. R., Int. J. Thermophys., 22, 227, 2001.
11. Frenkel, M.; Dong, Q.; Wilhoit, R. C.; Hall, K. R., Int. J. Thermophys., 22, 215, 2001.
12. TRC Thermodynamic Tables – Hydrocarbons, Thermodynamics Research Center: National Institute of Standards and Technology, Boulder, CO, 1942-2001.
13. TRC Thermodynamic Tables – Non-Hydrocarbons, Thermodynamics Research Center: National Institute of Standards and Technology, Boulder, CO, 1942-2001.
14. International Data Series-Selected Data on Mixtures, Thermodynamics Research Center: National Institute of Standards and Technology, Boulder, CO, 1979-2001.
15. <http://www.trc.nist.gov/>
16. Mandell, J., The Statistical Analysis of Experimental Data, Dover, New York, 1984.
17. Caracotsios, M.; Stewart, W. E., Introduction to Numerical Methods for Chemical Engineers, Stewart and Associates Engineering Software, Inc., 2001.
18. Bevington, P. R.; Robinson, D.K., Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, Inc., New York, 1992.
19. Press, W. J.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. R., Numerical Recipes in C, Cambridge University Press, Cambridge, 1988.
20. Gess, M. A.; Danner, R. P.; Nagvekar, M., Thermodynamic Analysis of Vapor-Liquid Equilibria: Recommended Models and a Standard Data Base, AIChE, New York, 1991.
21. Gmehling, J.; Onken, U.; Arlt, W., Vapor-Liquid Equilibrium Data Collection : Aqueous-Organic Systems, DECHEMA, Berlin, 1991.