The Dortmund Data Bank: A Computerized System for Retrieval, Correlation, and Prediction of Thermodynamic Properties of Mixtures[1]

U. Onken,[2] **J. Rarey-Nies,**[2] **and J. Gmehling**[2]

The Dortmund Data Bank (DDB) was started in 1973 with the intention to employ the vast store of vapor–liquid equilibrium (VLE) data from the literature for the development of models for the prediction of VLE. From the beginning, the structure of the DDB has been organized in such a way that it was possible to take advantage of the full potential of electronic computers. With the experience gained in fitting and processing VLE data, we extended the DDB system to other types of mixture properties, i.e., liquid–liquid equilibria (LLE), gas solubilities (GLE), activity coefficients at infinite dilution (γ^∞), heats of mixing (h^E), and excess heat capacities. Besides the files for mixture properties, the DDB contains pure-component data and program packages for various applications. New experimental data are checked for consistency before they are stored. For data retrieval user-specified search masks can be used. The data files are available via an online data service and through the Dechema Chemistry Data Series. For the purpose of data correlation and model testing, parameter fitting is performed with an optimization routine (Nelder–Mead). In the past years the DDB system has been successfully employed for the development of prediction methods for VLE, LLE, GLE, γ^∞, and h^E (UNIFAC, mod. UNIFAC, etc.).

KEY WORDS: activity coefficients (at infinite dilution); data bank; heats of mixing; mixtures; NRTL; phase equilibria; UNIQUAC; vapor–liquid equilibria; Wilson equation.

1. INTRODUCTION

Data on mixtures are needed for many chemical engineering calculations. These data are required especially for the design of various processes

---

for the separation of fluid mixtures from chemical reactions. Here, in particular, data on phase equilibria are of interest. Since distillation is the most frequently used separation process, vapor–liquid equilibria are predominantly important. Unfortunately experimental data on mixtures to be handled in such processes very often are not available. In principle, appropriate measurements are not difficult to perform, but they can be very time-consuming, especially for multicomponent systems. Although the amount of experimental work necessary to describe multicomponent systems can be considerably reduced by using models based on the concept of local composition, such as Wilson, NRTL, or UNIQUAC, equilibrium measurements of the relevant binary systems are often not feasible because of lack of time, for instance, when process alternatives have to be evaluated, as happens in process development and preliminary design. In these cases the parameters for binary systems for which no experimental data are available must be predicted. Regardless of the prediction method, the results will be more reliable if data for related systems have been used.

        With the idea to employ the vast store of published vapor–liquid equilibrium (VLE) data for the prediction of binary parameters for liquid mixtures, we started building up a computerized data bank of binary and multicomponent VLE data in 1973. Since our work on prediction methods had been aimed at nonelectrolyte systems, the data bank, now called Dortmund Data Bank (DDB) [1–4], is restricted mainly to mixtures of organic compounds but does include water. Pure-component data required for the evaluation and application of mixture data are, of course, also filed in the DDB. Subsequently the original data bank for VLE data was extended to other types of phase equilibria. Presently the DDB contains nearly all published experimental results on vapor–liquid equilibria, heats of mixing, activity coefficients at infinite dilution, and liquid–liquid equilibria for nonelectrolyte systems; furthermore, data banks for gas solubilities and excess heat capacities are being built up (cf. Table I) and a data bank for azeotropic data is planned.

        In addition to these data on mixtures, pure-component properties have been stored. The Dortmund Data Bank (DDB) also contains lists of references as well as a large number of programs, especially for data testing and retrieval, parameter optimization for various models, and process simulation.

        The present status of the DDB is shown in Table I. The main purpose when the DDB was started in 1973 was the development of a group contribution method. This goal was reached with the development of the widely used UNIFAC method [5]. With the data bank it is possible to test new modifications of prediction methods, other g^E models, and equations of state. Because of its structure the data bank can be directly integrated into process simulators.

Table I. Present Status of the Dortmund Data Bank

Pure-component properties	Approx. 2200 components	
Data on mixtures	Since	No. of isotherms or isobars[a]
Vapor–liquid equilibria	1973	12,800
Liquid–liquid equilibria	1977	3,000 (1)
Heats of mixing	1980	5,500 (1)
γ^∞	1984	22,000 values (2, 3)
Gas solubilities	1985	4,200 (4)
c_P^E	1986	400
Azeotropic data	1988	Planned
Lists of references	Approx. 6000 references from approx. 550 journals	

[a] In collaboration with (1) Instituttet for Kemiteknik, Danmarks Tekniske Højskole, Lyngby, Denmark [6, 7]; (2) Faculdade de Engenharia, Universidade do Porto, Porto, Portugal [8]; (3) Instituto di Chimica Applicata e Industriale, Universita di Trieste, Trieste, Italy [8]; (4) Technische Universität Berlin, Berlin, F.R.G.

2. STRUCTURE OF THE DATA FILES

Several records are used to store a VLE data set. The first record contains information on the data set. Therein the following information is stored:

   number of components,
   code number for data type,
   number of data points;
   system temperature or system pressure,
   code number for reference, and
   code numbers for the individual components.

Using the code numbers in the information record, any required property for calculating and representing this binary system may then be read from direct access files, whereby the key number is identical to the record number in the corresponding file. For each component the following data are stored:

   the compound name with the empirical formula,
   the constants A, B, and C of the Antoine vapor pressure equation for low and high pressure with range of validity T_{min}, T_{max},
   the molecular weight MW,
   the density ρ at temperature T_ρ, and
   the dipole moment μ.

Furthermore, the structural parameters r_i and q_i (for the UNIQUAC equation), the critical properties T_c, P_c, and v_c, the acentric factor ω, the melting temperature T_m, and the heat of fusion $\Delta_{fus}H$ are stored.

For the purpose of calculating activity coefficients with the UNIFAC method [5], the records for the pure components contain information on the molecular structure. Using the code number for the reference the program will find the complete original reference in the reference file.

3. HARDWARE CONFIGURATION

The complete data bank, including all auxiliary application programs, is implemented on a HP 1000/A900 minicomputer with 3-MByte internal memory and 700 MByte of fixed disk storage. This computer was chosen because of its highly transparent documentation and high-speed floating point calculations. The performance is about 50% that of a VAX 780. In addition for the HP-1000 with fixed disks, magnetic tape drive, plotter, and laser printer, there are various microcomputers which may act as graphic terminals or independent units for word processing, CAD, or other special applications. The data bank and the different program packages are going to be ported to run under the MS-DOS operating system and thus will be available on microcomputer in the near-future.

4. DATA RETRIEVAL AND PROCESSING
IN THE CASE OF VLE DATA

The first step is always finding the code numbers of the components of interest. This can be done using a program which searches the pure-component data file and finds all components with a specific empirical formula, name, or Chemical Abstracts Registring Number. An extended version of this program is available to find components with properties inside a specific range (e.g., vapor pressure, critical data, molecular weight, etc.).

Using the code numbers for the components, a selection program searches the mixture data file for all data on mixtures of any component contained in one list with any component from a second list. It is also possible to generate lists of all components containing a definite structural group, e.g., all alkanols or ketones.

The data output can be performed in various ways. The standard procedure for VLE, which is also used in publishing the VLE Data Collection [9], is as follows.

    (a)   The thermodynamic consistency of the data is checked with two different tests. The first test is performed by fitting the x, P, T

data to a flexible expression for the excess Gibbs energy. With the aid of the fitted parameters, the vapor composition y for each experimental point (x, P, T) in the data set is calculated and compared with the experimental value. If the mean deviation between experimental and calculated y values of a data set is smaller than 0.01, the data are considered to be consistent [10]. The second test uses the integrated form of the Gibbs–Duhem equation as proposed by Redlich and Kister [11]. For isobaric data the influence of the excess enthalpy is taken into account by an empirical parameter introduced by Herington [12]. Both tests can be used with complete data (x, y, P, T); the second test also works with x, y, T data.

(b) The parameters for five commonly used g^E models (Margules, Van Laar, Wilson, NRTL, UNIQUAC) are fitted to the experimental data. As optimizing strategy the program uses the SIMPLEX method according to Nelder and Mead [13]. Different objective functions are available $[\sum(\Delta\gamma_{rel})^2, \sum|\Delta y_{abs}|, \sum(\Delta P_{abs})^2, \sum|\Delta K_{abs}|)$.

(c) Using the model exhibiting the smallest mean deviation from experimental data in vapor-phase composition, an x, y diagram is generated. In addition, the mixture is tested for the existence of a miscibility gap. If an immiscible region is predicted, the compositions of the two liquid phases are calculated. This test is performed only for the "best" model, with the exception of the Wilson equation.

In general the vapor phase is treated as an ideal mixture. In case the mixture contains one or more carboxylic acids, vapor-phase nonideality is automatically taken into account using the chemical theory [14, 15]. As an option the vapor-phase nonideality may be calculated using the Soave–Redlich–Kwong equation of state [16] with k_{ij} being fitted or supplied by the user.

To examine data sets more closely it is advisable to inspect graphical representations on either the terminal or a plotter. A g^E model has to be specified together with an objective function to be used for the correlation. The parameters of this model are fitted to the data and used to calculate the curves in the plots.

Figure 1 presents diagrams obtained in this way for the vapor–liquid equilibrium data of the system benzene/hexafluorobenzene at a constant pressure of 760 mm Hg published by Chinikamala et al. [17]. Curves in Fig. 1 have been calculated with the NRTL equation. In this case deviations between experimental and calculated data are apparent for g^E at
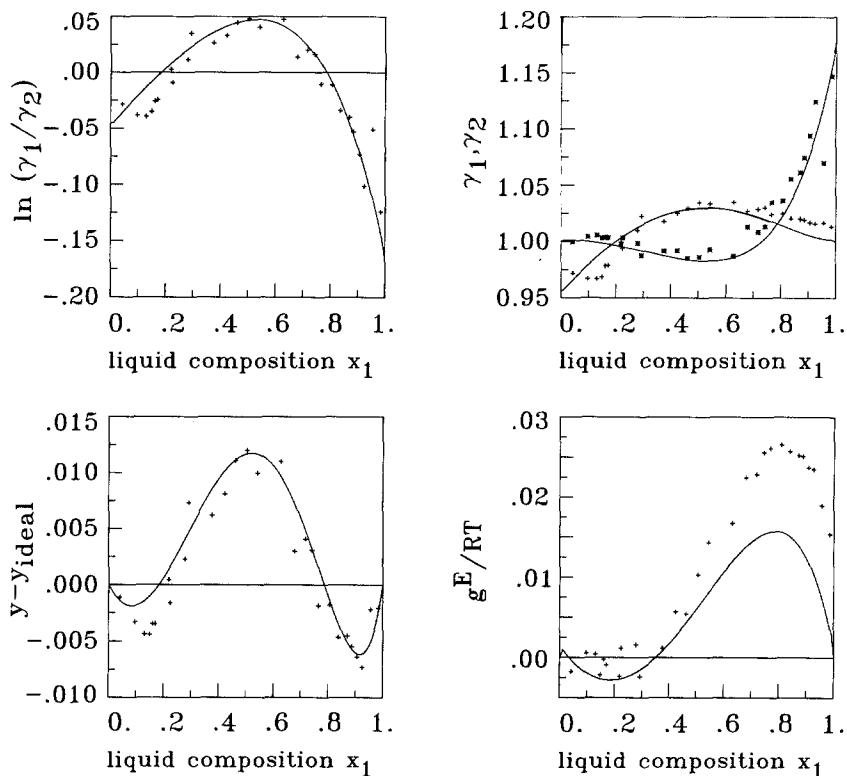
Fig. 1. Representation of vapor–liquid equilibrium data with different functions. System benzene/hexafluorobenzene at 760-mm Hg constant pressure.

higher concentrations of benzene, whereas for $\ln(\gamma_1/\gamma_2)$ and $y - y_{id}$ agreement seems to be rather good. It is noteworthy that in fitting the parameters of the NRTL equation, $\sum(\Delta\gamma_{rel})^2$ has been used as objective function. The reason for the systematic deviations in g^E for the data set in Fig. 1 is the limited flexibility of model equations with two or three parameters.

The procedures outlined above are restricted to processing a single data set. To fit a number of data sets simultaneously, a different set of programs is used. A data file is first generated using the selection program which contains not only the data set code numbers but the information record and the data records. This file is then processed by the consistency tests. With the help of an input file generator for the fitting program, one may select a subset of data sets from the data file. If different kinds of incomplete data $(x, y, T$ or $x, P, T)$ are included, it is not possible to use

the deviation in γ, y, or P as an objective function. In these cases a combined objective function has to be used. The weighting factors are supplied by default setting but may be modified interactively. In addition to VLE data, γ^∞ and h^E data may be used for fitting the parameters. The results of the simultaneous correlation can be examined using a graphical representation of the deviation of the data points versus liquid-phase composition. It is also possible to discard single data points or data sets and thus generate a new, more unique set of data for another parameter fit. This procedure can be used to compare different data sets at different temperatures or pressures, thus supplying additional information on the consistency of the data.

An example of using h^E data in addition to available VLE data for the representation of the vapor–liquid equilibrium of a binary system is shown in Fig. 2, where the temperature dependence of the azeotropic concentration of the system ethanol/water is plotted. Modeling this system with the Wilson equation yields a wrong tendency for the change of azeotropic composition with temperature. Only by including data on heats of mixing for the system and temperature-dependent interaction parameters can the correct temperature function for the azeotrope be obtained. It should be added that the quality of data representation for this system was essentially improved by taking into account nonideal vapor-phase behavior [18, 19].
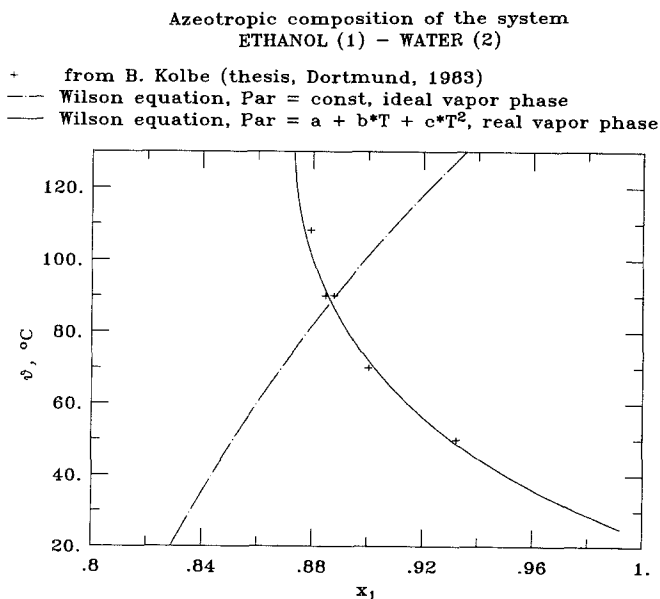
Azeotropic composition of the system
ETHANOL (1) − WATER (2)

+   from B. Kolbe (thesis, Dortmund, 1983)
— ·— Wilson equation, Par = const, ideal vapor phase
——   Wilson equation, Par = a + b*T + c*T², real vapor phase

Fig. 2. Dependence of the azeotropic composition on temperature in the system ethanol/water.

5. CURRENT APPLICATIONS OF THE DDB

Current applications of the DDB are mainly the

(a) development of prediction methods;

(b) testing and improvement of equations of state, group-contribution equations of state, and $g^E(\gamma)$ models; and

(c) integration in process simulators.

The stored data are also available via an on-line service provided by FIZ Chemie. In addition, the data are published via DECHEMA Chemistry Data Series [6-9], where the user can find additional information, such as model parameters, results of consistency tests, and comparisons between experimental data and values calculated from the various models as well as graphical representations of the data.

6. CONCLUSION

The Dortmund Data Bank for equilibrium data of mixtures containing organic compounds is stored on a Hewlett-Packard HP1000/A900 mini-computer together with a large number of application and service programs. It has been organized in such a way that it is easy to retrieve desired data, search the files for information-matching user-specified search masks, integrate the data bank into other computer programs, and update, edit, and test the data compilation.

ACKNOWLEDGMENT

REFERENCES

1. U. Onken and J. Gmehling, *Chem. Ing. Tech.* **49**:404 (1977).
2. U. Onken, *Chem. Ing. Tech.* **50**:760 (1978).
3. J. Gmehling and U. Onken, Proceedings of the Sixth International CODATA Conference (1978), p. 163.
4. J. Gmehling, CODATA-Bulletin 58, Thermodynamic Databases (Nov. 1985).
5. A. Fredenslund, J. Gmehling, and P. Rasmussen, *Vapor-Liquid-Equilibria Using UNIFAC* (Elsevier, Amsterdam, 1977).
6. J. M. Sørensen and W. Arlt, *Liquid-Liquid Equilibrium Data Collection, Vol. V* (DECHEMA Chemistry Data Series, Frankfurt, 1979).
7. C. Christensen, J. Gmehling, P. Rasmussen, and U. Weidlich, *Heats of Mixing Data Collection, Vol. III* (DECHEMA Chemistry Data Series, Frankfurt, 1984).

8. D. Tiegs, J. Gmehling, J. Bastos, A. G. Medina, M. Soares, P. Alessi, and I. Kicić, γ^∞-*Data Collection, Vol. IX* (DECHEMA Chemistry Data Series, Frankfurt, 1986).

9. J. Gmehling, U. Onken, W. Arlt, P. Grenzheuser, B. Kolbe, U. Weidlich, and J. R. Rarey-Nies, *Vapor-Liquid-Equilibrium Data Collection, Vol. I* in preparation (DECHEMA Chemistry Data Series, Frankfurt, 1977–1984).

10. H. C. Van Ness, S. M. Byer, and R. E. Gibbs, *AIChE J.* **19**:238 (1973).

11. O. Redlich and A. T. Kister, *Ind. Eng. Chem.* **40**:345 (1948).

12. E. F. G. Herington, *J. Inst. Petrol.* **37**:457 (1951).

13. J. A. Nelder and R. Mead, *Computer J.* **7**:308 (1965).

14. J. Marek and G. Standart, *Coll. Czech. Chem. Commun.* **19**:1074 (1954).

15. J. Marek and G. Standart, *Coll. Czech. Chem. Commun.* **20**:1490 (1955).

16. G. Soave, *Chem. Eng. Sci.* **27**:119 (1972).

17. A. Chinikamala, G. N. Houth, and Z. L. Taylor, *J. Chem. Eng. Data* **18**:322 (1973).

18. B. Kolbe and J. Gmehling, *Fluid Phase Equil.* **23**:227 (1985).

19. A. Gaulhofer, B. Kolbe, and J. Gmehling, *Fluid Phase Equil.* **39**:193 (1988).