

Factual data banks and their application to the synthesis and design of chemical processes and the development and testing of thermophysical property estimation methods*

Jürgen Rarey^{1,2,3,‡} and Jürgen Gmehling^{1,2}

¹Industrial Chemistry, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany; ²DDBST GmbH, Marie-Curie-Str. 10, 26129 Oldenburg, Germany; ³Thermodynamics Research Unit, University of KwaZulu-Natal, Durban, South Africa

Abstract: Factual data banks nowadays play an important role as a source for thermophysical property data for use in chemical process simulation, environmental models, and many other computer-based applications. In this work, the historical developments leading to modern factual data banks, the differences compared to other more bibliographically oriented data banks, and their most important applications and future potential will be discussed by using the example of the Dortmund Data Bank (DDB).

As the development of the different predictive models for mixtures is covered in a separate publication, this paper focuses on pure-component property estimation, regression of model parameters, test and verification of model parameters prior to process simulation, and advanced topics in process synthesis such as selection of entrainers and data-mining applications.

Keywords: data banks; group contribution methods; physical properties; process simulation; property estimation; thermodynamics.

INTRODUCTION

Synthesis, design, and simulation of chemical processes, especially thermal separation processes, are nowadays mostly carried out using sophisticated commercial process simulation software like Aspen Plus[®], ChemCAD[®], Hysys[®], ProII[®], Prosim[®], System/7[®].... The reliability and correctness of the simulation results is mainly influenced by the reliability and correctness of the thermophysical property parameters for pure components and mixtures. These should be verified using modern thermophysical property data banks and estimation methods prior to process simulation. In many cases, thermophysical property issues lead to problems in design and simulation [1].

Creative application of the thermophysical knowledge hidden in the vast amount of data and the parameter matrices of estimation methods can result in the detection of feasible and elegant process alternatives (computer-aided process synthesis and design) [2–5]. In addition, factual data banks and sophisticated software tools are ideal for the daily work of the chemical engineer, since they allow rapid

*Paper based on a presentation at the 20th International Conference on Chemical Thermodynamics (ICCT 20), 3–8 August 2008, Warsaw, Poland. Other presentations are published in this issue, pp. 1719–1959.

‡Corresponding author: E-mail: rarey@ddbst.com

retrieval, visualization, verification, and fitting of recommended values and parameters for thermodynamic property models [6].

In the case of optimal solvent selection for extractive or azeotropic distillation and liquid–liquid extraction, the automatic evaluation of a large number of entrainers is required, either by direct access to comprehensive factual data banks or through fast and reliable thermophysical property estimation methods. Both are possible due to the availability of large electronic data banks and integrated software packages.

The purpose of this paper is to describe the developments leading to modern factual data banks for thermophysical properties and highlight a number of important applications. Due to the authors' close familiarity and involvement with the Dortmund Data Bank (DDB) [7], this specific database will serve as an example throughout this publication. Yet, many or even most of the important features and specific design considerations exemplified by hand of the DDB are also true for other factual data banks.

While the multitude of applications of factual data banks in chemistry and chemical engineering is well accepted and acknowledged today, they were not immediately apparent when work on data compilations started. To understand the differences in design and focus between different data bank projects, it is therefore of interest to examine the historical boundary conditions and motivations of the time in which the projects were started.

HISTORICAL DEVELOPMENT

Probably the first large projects that later developed into important data banks were the different abstracts services. These allowed the reader to follow the research reported in a larger number of journals without the need to study each original publication. Abstracts were first included in primary scientific journals to inform the reader about research published in other journals. Due to the growing importance, journals especially devoted to abstracting appeared especially in the early 19th century, one of the most famous being the German *Chemisches Zentralblatt* (starting in 1830). Its primary focus was always on the German chemical literature, therefore, further abstracting journals were founded in other countries to specifically cover non-German publications. Today's largest abstracting service (Chemical Abstracts) appeared first in the United States in 1907.

Having abstracts of a large number of publications from different primary sources in one journal led to the construction of indexes, by which it was possible to conveniently locate relevant research in current or past publications. Abstract journals therefore proved to be of extraordinary scientific value. Chemical Abstracts was later computerized and now contains more than 26 000 000 abstracts and has a chemical substance index comprising more than 37 000 000 components.

The downside of abstracts lies in the fact that they only allow one to locate relevant information and to some extent judge the contents of an article. The information itself still has to be extracted from the original article. For this reason, practically all journals allow free access to the abstracts of published articles, which are usually provided by the authors together with the manuscript.

Another important step in the development of modern chemical and engineering data banks can be found in handbooks like those of Beilstein (1881), Landolt-Börnstein (1883), Gmelin (1887), In contrast to the abstract journals, these handbooks contain the most relevant information, so that in many cases access to the primary publication is no longer required. The information in handbooks is usually organized by chemical components instead of primary literature source. The Beilstein data bank today contains about 9 000 000 components, 10 000 000 reactions, and more than 320 000 000 property data (including molecular weights, melting points, ...). Beilstein and Gmelin are of especially great importance for organic respective of inorganic chemists, but have only found very limited use in chemical engineering applications. This is astonishing given the enormous amount of primary experimental data available through these handbooks. The reason is probably that the majority of the data covers components of little or no technical importance and the data are extracted and compiled by organic chemists

with limited knowledge of the specific requirements of process design and simulation. Beilstein employed various approaches to verify the consistency of their data like neural networks or quantitative structure–property relationships (QSPRs), but unluckily did not choose models and methods generally accepted in chemical engineering thermodynamics.

During the World War II, the importance of precise knowledge about the behavior of chemical components and their mixtures became apparent, especially with respect to hydrocarbons. Large amounts of fuel were required for transportation, ships, tanks, and warplanes. Chemical engineers required readily available and reliable data, especially PvT and enthalpy data of pure compounds and phase equilibrium data of mixtures. These were supplied in the form of data tables or, to avoid further computational effort, in the form of diagrams. Especially new technical developments like that of the nuclear bomb depended heavily on the availability of data as described by D. A. Palmer in his *Handbook of Chemical Engineering Thermodynamics* [8]:

II. IMPORTANCE OF PHYSICAL PROPERTIES

It sometimes seems that national security must be threatened to obtain concerted effort by a democratic nation. It was primarily the progress of science that finally ended the war with Japan. An almost forgotten element of the development of the atomic bomb relates to the race between the Allies and Germany to bring it to fruition. There were several ways to develop such a bomb. The short way required knowledge of a physical property of graphite. The Germans guessed at the answer, and guessed incorrectly. As a result they were led on the longer path of development of heavy water in Norway. The Americans knew of the German guess, but measured the physical property. They found that it was substantially different from the guess of the Germans, and were able to devise a much faster route to production of a nuclear weapon.^{1,9} If this scenario had been reversed, the U.S. might have become a province of the Third Reich!

1 National Research Council, National Needs for Critically Evaluated Physical and Chemical Data, Washington, D.C., 1978.

9 Sengers, J.V. and Klein M., Ed., Technological Importance of Accurate Thermophysical Property Information, Spec. Publ. 590, National Bureau of Standards, Washington, D.C.

Two of the most well-known producers for these tables were the Thermodynamic Research Center (TRC) founded in the United States in 1942 and the National Physics Laboratory (NPL) in the United Kingdom (ca. 1945). Due to the importance of reliable data in large-scale industrialization during the second half of the 20th century, similar efforts were started in many other countries and inside most chemical and petrochemical companies.

Today, knowledge of reliable thermophysical properties may not decide the outcome of a war, but nevertheless can have a significant effect in that one may

- find a better entrainer for a separation process,
- be able to use a lower number of stages or lower reflux ratio due to more precise phase equilibrium data used for distillation column design,
- use a specific solvent to improve equilibrium conversion in a chemical reaction, or
-

Overall, the availability of more precise and reliable data leads to better industrial performance using fewer resources and inflicting less damage upon the environment.

The way to modern thermophysical data banks was driven by different developments in the 1960s and 1970s of the last century:

- Algorithms and software for the simulation of unit operations, especially for distillation were developed and employed in chemical and petrochemical companies (Wang-Henke, 1966 [9], Naphthali-Sandholm, 1971 [10]).
- Thermodynamic models for the prediction of multicomponent mixture behavior from binary data (Wilson, 1964 [11], NRTL, 1968 [12], universal quasi-chemical theory (UNIQUAC), 1975 [13], SRK, 1972 [14], PR, 1976 [15]) became available.
- Sufficiently powerful computers allowed running more complex calculations.

The need to supply simulation software with the urgently required reliable thermophysical properties spurred great additional interest in experimental thermophysical property data. Data bank projects that originated around 1970 were consequently not primarily focused on providing human engineers with the urgently needed information. Instead, data were stored in a strictly standardized and computer-readable form to serve as input for model parameter regression with special focus on parameter determination and the development of predictive methods.

Data banks of this type initiated during this period in Germany included the Dortmund Bata Bank (J. Gmehling, U. Onken, 1973) for vapor–liquid equilibrium (VLE) data of normal and high boiling components, later followed by, e.g., the Berlin Data Bank (H. Knapp) for phase equilibria of low boiling components and DETHERM (1978, [16]), a project to merge different data banks into a unified form. Out of these efforts, DDB survived as an active producer while DETHERM serves as one of the data bank platforms to provide data from the DDB, ELDAR [17], and several other, mostly discontinued data banks to the different users in industry and academia. The current status and applications of the DDB will be given in more detail later on. A further data bank regularly updated by FIZ CHEMIE, Berlin and available online since 2003 is Infotherm, which covers data from approximately 18 500 publications [18].

In the United Kingdom, the Physical Property Data Service (PPDS) originated from NPL in 1972 and is now under the ownership of the German TÜV Bavaria. PPDS focuses mostly on specific problems in hydrocarbon processing and modeling of pure-component properties.

In the United States, the large collection of data of the TRC was brought into computer-readable form, retaining most of the more complex table structures that made it more convenient to use by engineers than by parameter regression software. Currently efforts are under way at NIST to extend and update the TRC data bank and improve its usefulness for chemical process simulation.

In addition, many large chemical and petrochemical companies and engineering firms started and maintain their own in-house data banks with company and public data. Today, the company data are often integrated into the in-house versions of licensed commercial data banks.

FACTUAL DATA BANKS AND SIMULATION

The increased availability of computers had a profound effect on science and technology. Phenomena lacking a closed-form analytical solution are more or less inaccessible without computers. Many practical problems require the solution of large coupled systems of inhomogeneous differential equations (fluid dynamics, chemical kinetics, ...), solution of large nonlinear sets of equations (distillation, process flowsheets, ...), or the generation of different configurations of molecular systems using Monte Carlo or molecular dynamics simulation.

In contrast to analytical solutions, numerical methods for solving differential equations can be linked to current experimental data of the real process via online data acquisition in order to synchronize simulation with real world observations. This opens the door to model predictive control (MPC) [19], whereby the controller contains a simulation of the controlled segment of the world. MPC then allows investigation of different control strategies in virtual reality (simulation) to select the optimum before applying it to the real system. Even without the aim of an advanced control strategy, realistic simulation models combining

- mathematical models of the physical system (mass, energy, and impulse balances);
- thermodynamic models describing the behavior of the pure components and mixtures in the different phases as well as the equilibrium between the phases (separation factors* for, e.g., distillation, liquid–liquid extraction, supercritical extraction, ...);
- transport properties that describe viscosity, thermal conductivity, diffusion, ...; and
- ...

are of great interest. The interaction between the simulation model and the real world segment is shown in Fig. 1.

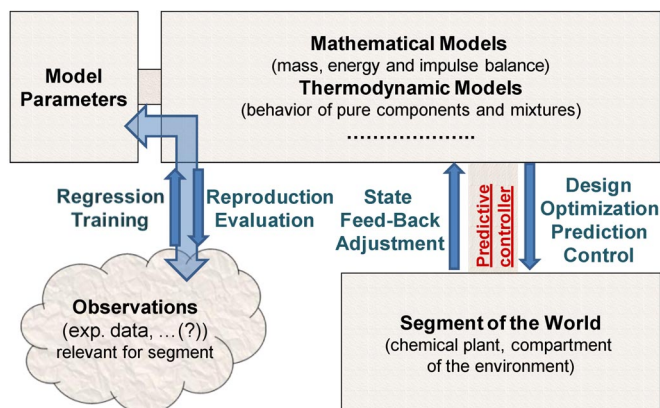


Fig. 1 Application and training of dynamic simulation models.

The huge importance of business simulation models led to the rapid success of the first personal computers, namely, the Apple® II running the spreadsheet calculation software VisiCalc® in 1979.

Since then, these models were introduced into many other fields so that now, in the first decade of the 3rd millennium, a large number of important engineering, economic, business, and political decisions are already strongly based on the results of simulation models (e.g., climate models for prediction of global warming, ...).

The results of these simulation models are strongly affected by the different assumptions and model parameters. That is why these parameters should be fitted to reliable data (observations) about simplified systems. Regressing parameters to observations about the complex target system is usually not feasible as these observations do not sufficiently isolate the influence of different model parameters and variables. This constitutes the great importance of experimental data for isolated subsystems that need to be used to “train” the various model parameters in the simulation model.

As the number of experimental data is limited and new data cannot be economically produced in very large quantity, estimation methods or even predictions from first principles are often considered as potential alternatives of lower reliability.

DEVELOPMENTS OF FACTUAL DATA BANKS, THE DORTMUND DATA BANK

In the early days (1960–1980), computer simulation of unit operations in chemical engineering was heavily focused on distillation. Distillation columns can be approximated by simple equilibrium stage

*In the case of VLE, the separation factor between two components i and j is defined as $\alpha_{i,j} = (y_i/x_i)/(y_j/x_j)$, where x denotes the liquid-phase mole fraction and y the vapor-phase mole fraction.

models, and the required thermophysical properties are predominantly the distribution factors of the components between liquid and vapor phase (VLE) and to some extent heat capacities and enthalpies of the liquid and vapor phase [9,10].

As data for multicomponent mixtures are usually not available, models had to be developed that allowed to predict the multicomponent mixture behavior based on data for the binary systems (Wilson [11], NRTL [12], UNIQUAC [13], SRK [14], PR [15]). It quickly became apparent that for many mixtures of practical interest, this information was not available and alternative routes via estimation methods were required. Models based on pure-component data alone like the regular solution theory of Scatchard and Hildebrand [20] proved to be inadequate for polar systems. The first successful model based on the solution of groups concept was the analytical solutions of groups (ASOG) method by Derr and Deal of Shell [21] published in 1969.

In 1973, J. Gmehling, at that time a young chemist, started work on a comparable method in the group of U. Onken at the University of Dortmund. U. Onken had previously been responsible for distillation design at the company HOECHST in Germany and was immediately aware of the importance of ASOG as an alternative route to urgently needed information about the binary and multicomponent mixture behavior. After teaming up with Aa. Fredenslund of Lyngby, Denmark, they decided to work together on the further development of the the universal functional activity coefficient (UNIFAC) method [22]. UNIFAC and ASOG are both based on the solution of group concept, but UNIFAC incorporated the UNIQUAC instead of the Wilson model.

This work led to the publication of a book on UNIFAC, which included the exact calculation procedure, a large number of group interaction parameters, numerous applications of UNIFAC, and extensive software source code in 1977 [23].

Development of such a method would not have been possible without the availability of a large amount of experimental phase equilibrium data (in case of UNIFAC, only VLE data were used). This was the reason that in 1973 work was started on an electronic data bank for VLE data of normal boiling components with support by H. Stage [24], who had already compiled a significant amount of data. Owing to great interest in this data compilation, data were published in form of a printed data series starting in 1977 [25]. Data representations in these books focused on the raw experimental data, and regression results for the most commonly used g^E models* as well as the results of two thermodynamic consistency test are given where applicable (Fig. 2).

This form of publication provided exactly the information required for modeling of industrial distillation processes. In contrary to other previous data bank projects, the stored information was more focused on the actual experimental data and the data bank entries were not aimed at replacing the original reference as source of complimentary information. Parallel to the publication in form of data tables, the DDB was delivered as an electronic in-house database.

As a result of the great success of the UNIFAC method, work was later extended to further types of data in order to develop an improved UNIFAC version (modified UNIFAC (Do) [26,27]) as well as further models for the description of high-pressure phase equilibria (PSRK [28,29], VTPR [30]), electrolyte mixtures (LIQUAC [31–34], LIFAC [35]), Table 1 summarizes the most important types of data required for the different estimation methods mentioned.

*The excess Gibbs energy is equal to the difference between the Gibbs energy of the real mixture and an ideal mixture (following Raoult's law) at the same temperature, pressure, and composition. The partial molar excess Gibbs energy of component i is directly connected to the activity coefficient of component i via the equation $\bar{g}_i^E = RT \ln \gamma_i$. Therefore, g^E -models provide a convenient route to the activity coefficients that are often used to represent the real behavior of liquid or solid mixtures.

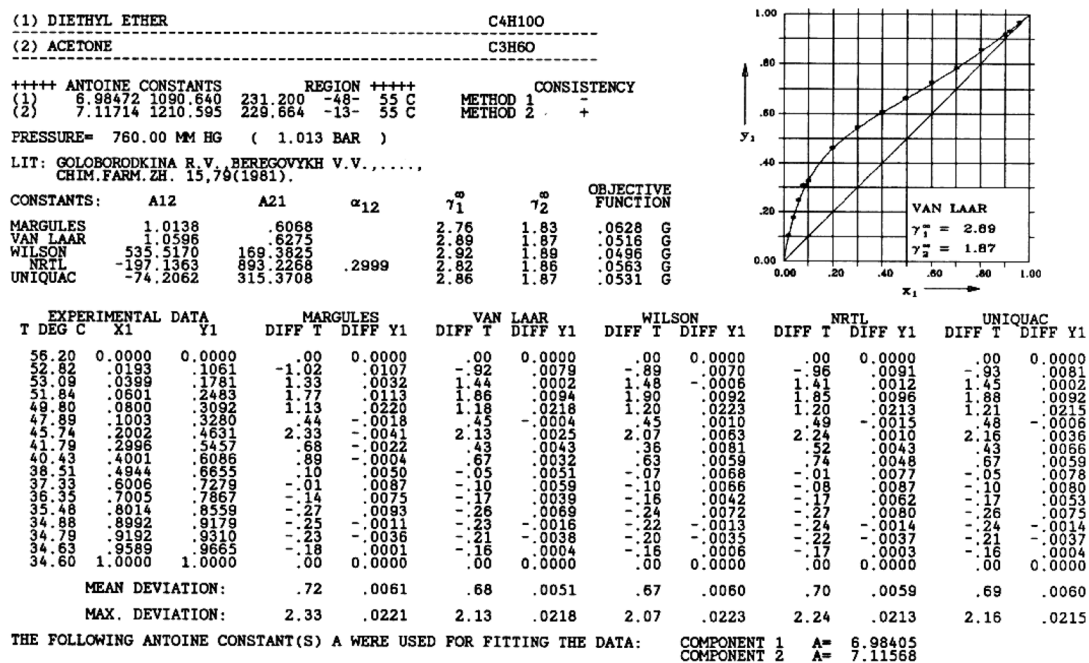


Fig. 2 VLE data representation as used in the DECHEMA Chemistry Data Series, Vapor-Liquid Equilibrium Data Collection [25].

Table 1 Data used for the development of some property estimation methods using the DDB.

Model	Required data	Abbrev.
Original UNIFAC [22,23,36]	Vapor-liquid equilibria of normal boiling components	VLE
Modified UNIFAC (Do.) [26,37]	Vapor-liquid equilibria of normal boiling components Excess enthalpies Excess heat capacities Activity coefficients at infinite dilution Liquid-liquid equilibria Solid-liquid equilibria Azeotropic data	VLE HE CPE ACT LLE SLE AZD
KOW-UNIFAC [38]	Octanol-water partition coefficients	KOW
Modified UNIFAC for freons [39]	As for modified UNIFAC	
Modified UNIFAC for ionic liquids [40]	As for modified UNIFAC	
PSRK [28,29], VTPR [30] (group contribution equations of state)	Vapor-liquid equilibria of low boiling components Gas solubilities Excess volumes and densities of mixtures Critical data of mixtures	HPV GLE VE CRI
LIQUAC [31-34], LIFAC [35] (electrolyte solution models)	Vapor-liquid equilibria and osmotic coefficients of electrolyte solutions Salt solubilities	ELE ELSE

(continues on next page)

Table 1 (Continued).

Model	Required data	Abbrev.
PSRK [41], VTPR (extension to electrolyte mixtures)	Gas solubility in electrolyte systems	EGL
COSMO-RS(OI) [42,43]	As for modified UNIFAC Molecular structures Quantum chemical (DFT) calculations	
Pure-component property estimation Models (T_b , T_c , P_c , V_c , $P^s(T)$, $\eta(T)$, ...)	Pure-component properties Normal boiling temperature (T_b) [44,45] Vapor–liquid critical data (T_c , P_c , V_c) [46] Saturated vapor pressure as function of temperature [$P^s(T)$] [47,48] Liquid saturated viscosity as function of temperature [$\eta(T)$] [49]	PURE
(Cooperation)	Various data of polymer systems	POLY

During the development of the DDB, several rules were strictly applied:

- Data were only entered from the primary publication.
- All primary publications are available as copies for data verification.
- Only experimental data were considered.
- Development of individual data banks was performed by specialists for the respective types of data, usually scientists who at the same time were in charge of model development.

Especially the intensive use of the data for the development of predictive methods led to the identification of a large number of errors in the primary literature (wrong components, wrong component order, wrong units, typographic errors, ...). Due to the success of the published Vapor–Liquid Equilibrium Data Collection, also further printed data collections were published within the DECHEMA Chemistry Data Series [50–52] and by VCH [53].

In 1989, a spin-off of the research group of Prof. Gmehling (University of Oldenburg) (DDBST GmbH, <www.ddbst.com>) took over the further development and update of the DDB. In addition to the different data banks, DDBST delivers software tools for data handling, retrieval, correlation, estimation, and visualization as well as process synthesis tools (DDB Software Package DDBSP).

The current status of the individual data banks in the DDB data pool is shown in Fig. 3.

As of 2009, the DDB contains experimental data from 64 300 references from approximately 1800 journals for more than 32 000 compounds including salts, adsorbents, and polymers. In addition, 25 700 molecular structures are stored in form of connectivity matrices and more than 48 700 sets of model parameters were derived by regression of vapor pressures, liquid densities, heat capacities, ... of pure components as well as simultaneous regression to VLE, h^E , γ^∞ , ... data for mixtures. For the use of modern methods employing results from quantum chemistry, more than 2840 COSMO σ -Profiles were derived using a high level of theory.

Besides data from the open literature, a significant portion of experimental data originates from private communications, M.Sc. and Ph.D. theses as well as from chemical industry.

DDB covers worldwide sources in all languages. 100 % of all costs are covered by commercialization, DDBST receives no public funding. Special conditions are available for the use of the DDB in research by individual research groups, whole departments or research institutions in a country. DDB data banks are available under different software platforms (DDBSP, DETHERM, ...), in-house, online, via data service or as printed data series by DECHEMA [25,50–52] and VCH [53].

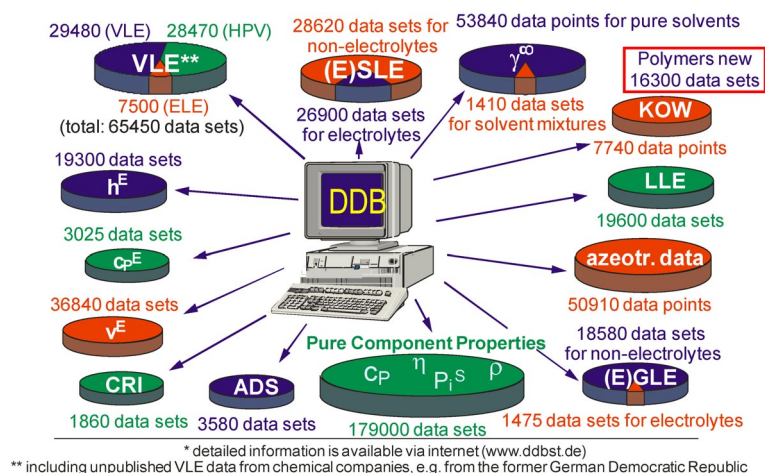


Fig. 3 Current status of the DDB (version 2009). Abbreviations: ELSE: solid–liquid equilibria of electrolyte mixtures; EGLE: gas solubility in solvents containing electrolytes; ADS: adsorption equilibria; γ^∞ : activity coefficients at infinite dilution. All other abbreviations are defined in Table 1.

DIFFERENT APPLICATIONS OF FACTUAL DATA BANKS BY EXAMPLE OF THE DDB

In order to gain the most benefit out of the significant investment in building up and maintaining a large factual data bank and to ensure maximum reliability of the data, work on the DDB is integrated into a network of research, production, education and sales/marketing efforts as shown in Fig. 4.

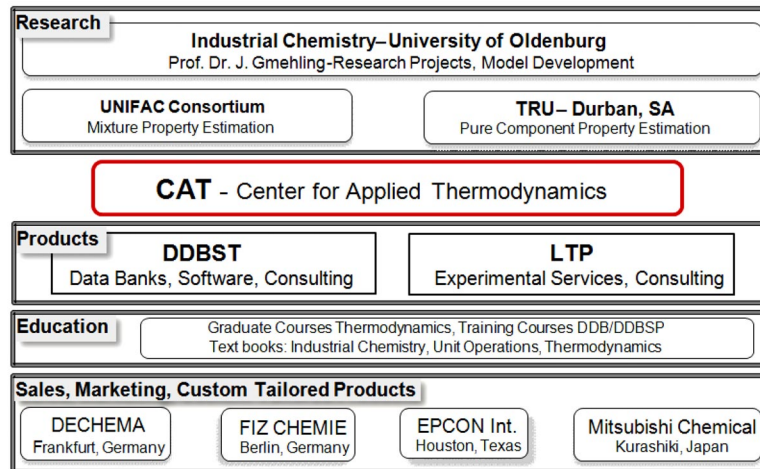


Fig. 4 DDB in the context of CAT (Center for Applied Thermodynamics).

The major part of the research stems from the research group at the University of Oldenburg (<www.uni-oldenburg.de/tchemie>) in case of VTPR and COSMO-RS(OI) and the consortium (<www.unifac.org>) for the further development of UNIFAC, modified UNIFAC, and PSRK [42]. Development of pure-component property estimation methods is conducted in cooperation with the Thermodynamic Research Unit (TRU) at the University of KwaZulu-Natal in Durban, South Africa.

Further cooperation exists with research groups in Beijing (models for electrolyte solutions), Lyngby (modeling of polymer systems),

DDBST builds and maintains the data banks and software packages and is responsible for marketing, sales, and customer support. In addition, DDBST is strongly engaged in user specific consulting services, partly in cooperation with its partner company LTP (experimental services, <www.ltp-oldenburg.de>).

In order to educate chemical engineers to be able to better exploit the possibilities of modern factual data banks, software, and estimation methods, regular graduate courses (typically, 3 to 7 per year) in applied thermodynamics for the synthesis, design, and simulation of thermal separation processes are carried out by the authors at various places world-wide*. These courses are supported by elaborate course material and various textbooks.

Besides DDBST, other companies and organizations (e.g., DECHEMA) provide in-house or on-line access to the DDB.

Since work on the DDB started 35 years ago, its primary mission has always been the development of reliable estimation methods for pure-component and mixture behavior. An overview of the work on mixture property estimation was given in the 2008 Rossini Lecture (ICCT 2008 in Warsaw) and was published separately elsewhere [54].

DEVELOPMENT OF ESTIMATION METHODS FOR PURE-COMPONENT PROPERTIES

Starting in 1991, the DDB was extended to different pure-component properties. Due to the vast amount of data published in literature, this data bank is now the largest single subfile of the DDB. The current status is shown in Table 2.

Table 2 Current status of the pure-component property database (April 2009, DDB).

	Components	References	Data Sets	Data Points
Vapor Pressure	8,624	8,169	31,214	197,891
Critical Data	1,100	997	3,694	3,703
Viscosity	3,403	3,157	19,072	116,049
Density	8,578	7,835	48,479	338,284
Melting Point	7,362	3,441	15,724	17,603
Heat Capacity	4,333	3,054	14,981	265,073
Enthalpy of Vaporization	2,799	1,200	5,601	14,754
Enthalpy of Fusion	1,905	1,069	2,856	3,006
Enthalpy of Sublimation	338	167	544	1,375
Thermal Conductivity	872	984	9,292	99,664
Surface Tension	2,246	765	5,143	23,373
Entropy (various)	1,650	1,092	3,036	16,978
Enthalpy of Combustion	1,730	543	2,002	2,011
Enthalpy of Formation	3,125	1,333	4,785	5,853
Speed of Sound	605	512	4,035	32,427
G-Function	1,092	564	1,770	26,574
...				
Total	19,183	22,696	175,590	1,187,174

As in the case of mixture data, the primary goal was to develop reliable estimation methods with a large range of applicability. In contrast to mixture data estimation, numerous models had already been published in literature. A major part of these methods employed the group contribution approach, often with a very complex set of structural groups and second-order corrections. This complexity allowed capturing more details of molecular structure and thus improved the estimation results, but at the same time it was also a major drawback since these complex methods were difficult to apply by hand. It was,

*<<http://www.ddbst.com/new/GVTCourses.htm>>

therefore, not astonishing that the most popular methods like that of Joback and Reid [55] employed only a small number of relatively simple groups.

To improve this situation and to avoid the tedious manual group fragmentation for a large number of components and different models, several important software tools were developed [56]:

- an editor program to draw and display molecular structures (ARTIST),
- a molecular structure data bank for all components in the DDB (ChemDB),
- an automated procedure to correctly identify the different structural groups in a molecule (AutoInkr), and
- separate group structure definitions for each group contribution methods defining the group and its valid context inside a molecule.

The program package ARTIST integrates these software tools and also offers estimation quality assessment based on the chemical family (ethers, ketones, ...) of a component, interfaces to commercial quantum chemical software, basic structure optimization, substructure search, ... Approximately 100 different group contribution methods are available for a large variety of different properties.

With the availability of a vast amount of experimental data and the required software to minimize tedious and both time-consuming and error-prone tasks, work started on the testing of literature models and the development and regular extension and revision of new pure-component property estimation methods for

- normal boiling temperature (T_b) [44,45];
- vapor-liquid critical data (T_c , P_c , V_c) [46];
- liquid-saturated vapor pressure as function of temperature [$P^S(T)$] [47,48]; and
- liquid-saturated viscosity as function of temperature [$\eta^S(T)$] [49].

Methods for surface tension [$\sigma(T)$] and thermal conductivity of liquids [$\lambda(T)$] are currently under development. Table 3 shows a typical comparison of different methods for the estimation of the normal boiling point. In this table, results for all components that a method could be applied to were used. This comparison of the methods is therefore not based on a common set of components.

Table 3 Typical comparison of different methods for the estimation of the normal boiling point for halogenated hydrocarbons (NR: Nannoolal et al. [45]; JR: Joback and Reid [55]; SB: Stein and Brown [57]; CG: Constantiou and Gani [58]; MP: Marrero and Pardillo [59]; CR: Cordes and Rarey [44]; EW: Ericksen et al. [60]) (taken from [45]).

Compounds	Absolute Average Deviations (K) (number of components as superscript)						
	NR	JR	SB	CG	MP	CR	EW
Fluorinated Saturated (HC)	6.37 ⁶⁴	18.53 ⁶⁴	59.89 ⁶⁴	25.62 ⁶⁴	14.25 ⁶³	7.35 ⁶⁴	5.5 ⁴¹
Fluorinated (HC)	7.16 ⁸⁶	18.01 ⁸⁶	48.96 ⁸⁶	26.46 ⁸⁶	12.83 ⁸³	8.02 ⁸⁶	6.53 ¹⁴
Chlorinated Saturated (HC)	7.53 ⁶⁴	26.01 ⁶⁴	14.72 ⁶⁴	8.26 ⁶⁰		7.43 ⁶⁴	9.72 ⁴⁹
Chlorinated (HC)	6.96 ¹¹⁷	21.36 ¹¹⁷	14.05 ¹¹⁷	9.36 ⁹⁵		6.62 ¹¹⁷	7.94 ⁹⁸
Brominated Saturated (HC)	5.48 ³¹	14.76 ³¹	8.82 ³¹	6.91 ³¹		6.34 ³¹	16.93 ²⁸
Brominated (HC)	5.8149	12.89 ⁴⁹	8.88 ⁴⁹	7.48 ⁴⁹		7.36 ⁴⁹	17.98 ⁴⁴
Iodinated(HC)	5.14 ¹⁸	13.63 ¹⁸	6.91 ¹⁸	5.77 ¹⁸	12.06 ¹⁵	5.20 ¹⁸	13.43 ¹⁷
Halogenated Compounds	8.09 ⁵⁶⁵	22.63 ⁵¹⁴	23.15 ⁵²⁰	18.86 ⁴³⁰	12.83 ¹³²	9.09 ⁵⁶⁴	20.69 ³⁸²

In order to assist the engineer to select the best method(s), the software automatically analyzes the structure of the molecule and determines the chemical families it belongs to (e.g., aliphatic alcohols, aromatic hydrocarbons, ...). The results of the method tests are then presented in the form of the typical deviations of each method for the respective chemical families.

Besides the development of new and superior estimation methods, this work improved the quality of the data bank as erroneous or questionable data could be identified. Similar work had been conducted, e.g., by the DIPPR 801 project as part of the software package DIADEM [61]. In contrast to ARTIST, which employs connectivity matrices and 3D coordinates to store and fragment molecular structures, DIADEM uses so-called SMILES strings which were a popular method to enter molecular structures into text-based computer software.

REGRESSION AND VERIFICATION OF PROPERTY MODEL PARAMETERS PRIOR TO PROCESS SIMULATION

Parameter regression and parameter data banks

As already shown in Fig. 1, training and verification of thermophysical property models is a major application of factual thermophysical property data banks.

The software package DDBSP provides numerous options for this purpose. It features dynamic data evaluation and parameter regression, interactive graphical software, and full documentation of included and discarded data. This allows easy upgrade by DDBST if new data become available or upgrade by customers employing their proprietary data. Besides the data regression capabilities integrated with the DDB, further regression packages are offered by numerous institutions and companies, e.g., DECHEMA (DPP [62] integrated with DETHERM) or the ASPEN Data Regression System (DRS).

Regression of primary experimental data can be a difficult task depending on the amount and quality of the available data, and for this reason, compilations of regressed parameters have been prepared by various authors within publicly or privately sponsored projects (e.g., DIPPR 801 [63]) or by private companies (DDBST, PPDS). The majority of these parameter compilations cover properties of pure components only.

The currently most well-known compilation prepared by the DIPPR 801 project covers a multitude of pure-component properties for a large set of organic components of industrial importance. It is mainly based on a source database of experimental data, but predictive correlations were employed whenever no sufficiently reliable data were available. Its basic mission is to provide complete coverage of all properties for a given list of components.

An analysis of the DIPPR correlations of second virial coefficients of pure fluids conducted in 1998 revealed that 25 % of the data used for these correlations were taken from primary references. 10 % came from monographs, 12 % from handbooks, and 53 % were generated by estimation methods. All values and parameter sets contain information about their probable reliability.

In 2008, DIPPR 801 provides 34 constant properties and 15 temperature-dependent properties for 1944 components (2069 components in the sponsor version). These parameters are distributed together with various software products, e.g., Aspen Properties, EPCON's API Technical Databook and Engineer's Aide, ... Recommended constant and temperature-dependent properties have been used for the development of various property estimation methods (QSPR). DIPPR 801 provides regression and visualization software, which also includes several recommended predictive methods.

Mixture parameters

While in the case of pure-component parameters and temperature-dependent properties, reliable sources for recommended parameters are available and users in many cases do not experience the necessity to verify their parameters against original experiments, the situation is strikingly different when it comes

to mixture parameters. Especially, binary interaction parameters for g^E -models or equations of state are of great importance for the correct description of the mixture behavior and the simulation of a unit operation. Parameter data banks for mixture parameters are often delivered with commercial software products like process simulators. Although significant care and effort was invested into deriving these parameters, they may easily produce erroneous results in specific situations. Simulator companies therefore generally recommend verifying the applicability especially in the case of binary interaction parameters.

As a first example, the descriptions of the VLE of the binary mixture acetone (1)–water (2) at 101.3 kPa using the NRTL model with binary interaction parameters provided with a simulation software are shown in Fig. 5.

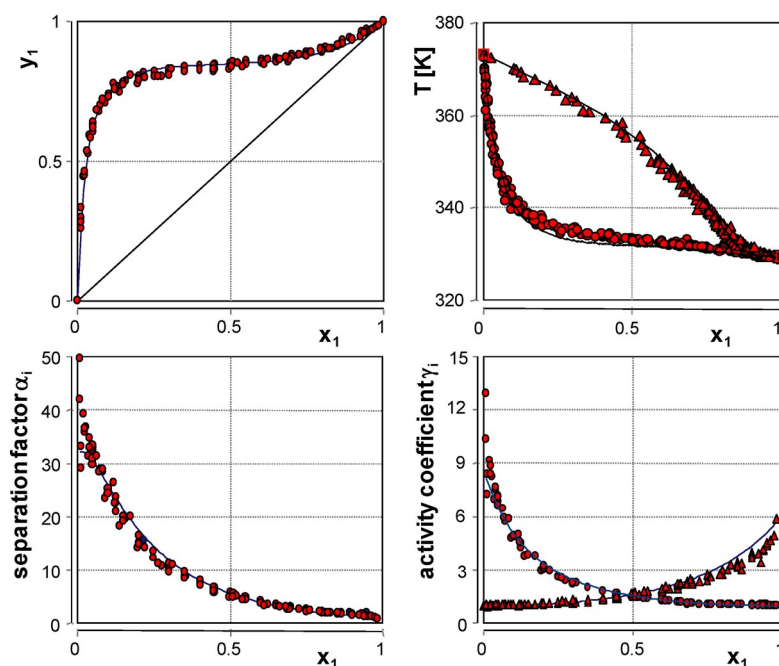


Fig. 5 Different graphical representations of the binary VLE behavior of the mixture acetone (1)–water (2) at 101.3 kPa together with curves calculated from NRTL using binary interaction parameters delivered with a process simulator.

At first glance, the regression seems to adequately reproduce the experimental data, which show a positive deviation from Raoult's law. Only at low acetone concentrations, the activity coefficient of acetone and the separation factor are significantly too low. This would mean that in order to diminish the acetone content of a liquid waste stream, the model calculation would result in a greater part of the stream to be evaporated.

A more typical industrial situation is the production of high-purity acetone in a distillation column, whereby the purity of the bottom (water) product is of less concern. In a mixture with positive deviation from Raoult's law, the highest separation effort is at the top of the column [64]. Figure 6 shows the separation factor between acetone and water in the concentration range above 90 mol % of acetone calculated using three different models (Wilson, NRTL, UNIQUAC). All parameters were supplied by the same simulator. The separation factor α_{ij} is an important quantity. In case α_{ij} is unity, the components i and j cannot be separated by ordinary distillation (azeotropic condition).

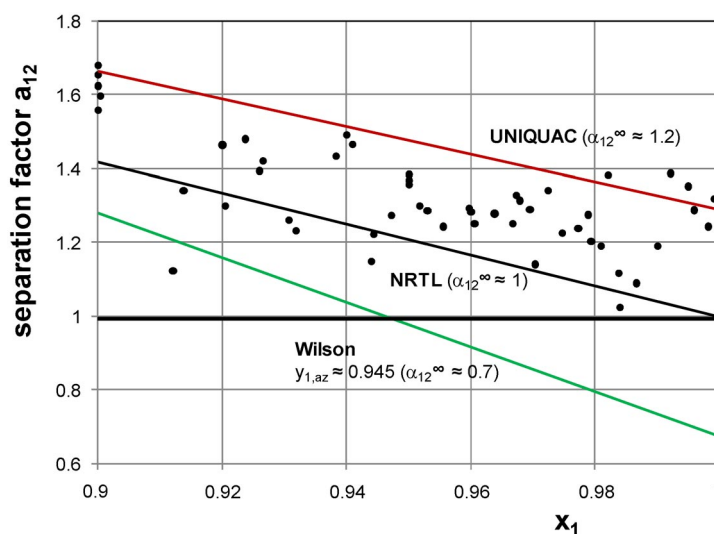


Fig. 6 Experimental separation factor α_{12} for the binary system acetone (1)–water (2) at 101.3 kPa together with curves calculated using Wilson, NRTL, and UNIQUAC with binary interaction parameters delivered with a process simulator.

As can be seen in Fig. 6, the Wilson model falsely describes an azeotrope at 94.5 mol %, NRTL results in a very difficult separation where α_{12} approaches unity, and UNIQUAC predicts the lowest separation effort. Judging from the data, the truth is probably somewhere between the NRTL and UNIQUAC results. Simultaneous regression to VLE and experimental activity coefficients at infinite dilution would lead to a reliable value of $\alpha_{12}^{\infty} = 1.16$.

Designing an acetone purification using NRTL with the parameters from the simulator would result in a very large column (large number of stages and/or high reflux ratio). Selecting UNIQUAC would seriously under-design the separation.

Most VLE data are usually available at a pressure at or close to 101.3 kPa, while distillation columns are often operated under elevated pressure to increase the capacity for a given equipment size and to improve heat integration. It is, therefore, important to analyze given parameters with respect to the correct temperature dependence of the real behavior. The temperature dependence of the activity coefficients is defined by the Gibbs–Helmholtz equation and given by the partial molar excess enthalpy \bar{h}_i^E :

$$\left(\frac{\partial \ln \gamma_i}{\partial (1/T)} \right)_{P,x} = \frac{\bar{h}_i^E}{R} \quad (1)$$

If the correct excess enthalpy is predicted by the model, an extrapolation to other temperatures is possible.

Figure 7 shows experimental excess enthalpy data from the DDB together with curves calculated from the very popular NRTL model using simulator parameters. The calculations are not even qualitatively correct.

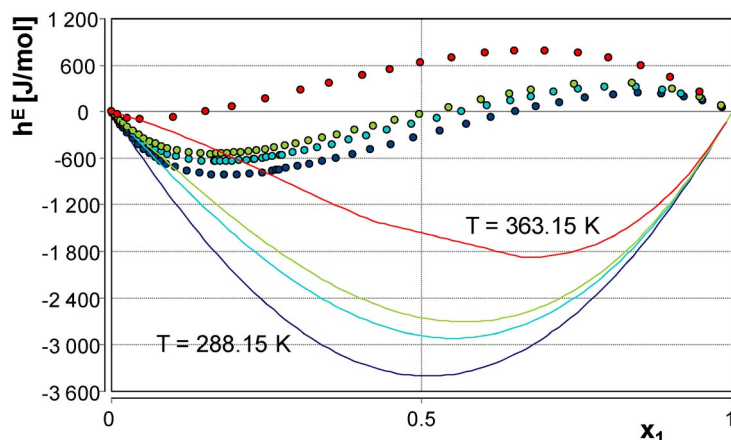


Fig. 7 Experimental excess enthalpy data for the binary mixture acetone (1)–water (2) from the DDB together with curves calculated from the NRTL model using simulator parameters.

Because the excess enthalpy data are not described correctly, erroneous prediction of the activity coefficients at infinite dilution and the azeotropic composition as a function of temperature are obtained as shown in Fig. 8. Figure 9 shows results for the NRTL model with binary parameters fitted simultaneously to VLE, activity coefficients at infinite dilution, and excess enthalpy data and azeotropic data using temperature-dependent interaction parameters to demonstrate the ability of NRTL to reproduce the experimental findings. DDBST is strongly engaged in the development of high-quality temperature-dependent interaction parameters for a large variety of binary mixtures.

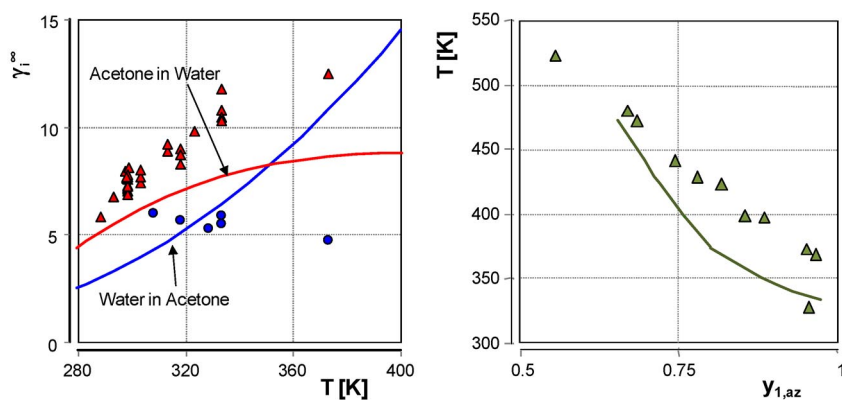


Fig. 8 Experimental activity coefficient at infinite dilution data and azeotropic compositions for the binary mixture acetone (1)–water (2) from the DDB together with curves calculated from the NRTL model using simulator parameters.

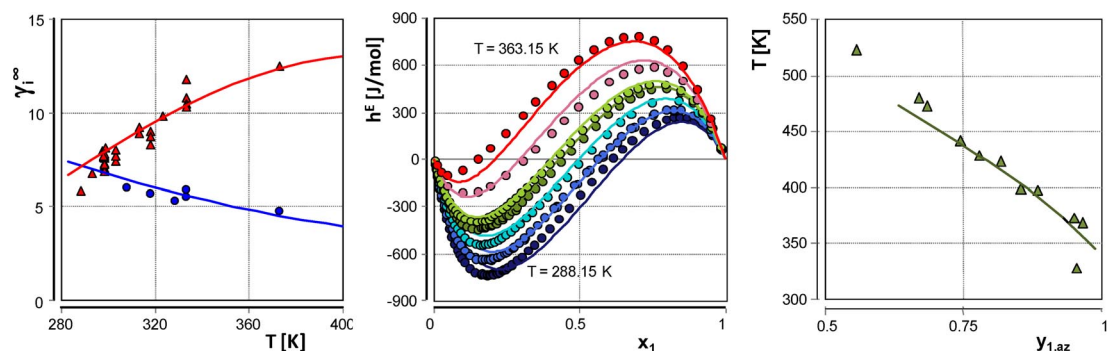


Fig. 9 Experimental excess enthalpies and activity coefficients at infinite dilution data and azeotropic compositions for the binary mixture acetone (1)–water (2) from the DDB together with curves calculated from the NRTL model using temperature-dependent binary interaction parameters.

In most practical applications, binary interaction parameters are used to predict the behavior of a multicomponent mixture. Especially in case of separations involving the VLE, correct prediction of all binary and higher azeotropic points is of great importance. In addition, no unrealistic azeotropic behavior should be falsely predicted. Table 4 shows a comparison between the predictions using NRTL with model parameters delivered from a simulator and experimental data from the DDB for a four-component system at 101.3 kPa. In this case, all calculation results match the azeotropic and zeotropic data from the DDB. In case the operating conditions deviate from ambient pressure, the correct description of all excess enthalpy data should also be verified.

Table 4 Experimental azeotropic compositions (mean values of data stored in the DDB [7]) for the system benzene (1)–cyclohexane (2)–acetone (3)–ethanol (4) at 101.3 kPa and results from the NRTL model using parameters delivered from the simulator.

system	calc. (NRTL – Simulator)				experimental*			
	type of azeotrope	$\vartheta / ^\circ\text{C}$	$Y_{1,\text{az}}$	$Y_{2,\text{az}}$	type of azeotrope	$\vartheta / ^\circ\text{C}$	$Y_{1,\text{az}}$	$Y_{2,\text{az}}$
1 - 2	homPmax	77.5	0.550		homPmax	77.6	0.543	
1 - 3	none				none			
1 - 4	homPmax	67.7	0.554		homPmax	67.9	0.552	
2 - 3	homPmax	53.5		0.233	homPmax	53.2		0.248
2 - 4	homPmax	64.9		0.559	homPmax	64.8		0.553
3 - 4	none				none			
1 - 2 - 3	none				none			
1 - 2 - 4	homPmax	64.7	0.116	0.461	homPmax	64.9	0.113	0.462
1 - 3 - 4	none				none			
2 - 3 - 4	none				none			
1 - 2 - 3 - 4	none				none			

In some cases, an inadequate description of one of the binary subsystems can result in catastrophic effects on the results for the multicomponent mixture. This is especially true in ternary mixtures consisting of two components (1) and (2) forming an azeotropic or near-azeotropic mixture and an extractive solvent (3). As the extractive solvent usually has a significantly higher boiling temperature, adding larger amounts of entrainer will strongly increase the system temperature and require binary interaction parameters for the (1)–(2) mixture that are valid at the elevated temperature. At the same time, the ratio of the activity coefficients of (1) in (3) and (2) in (3) need to be correctly described

at higher solvent concentrations as this determines the effect of the selective solvent on the separation of the (1)–(2) mixture by extractive distillation. VLE data for the (1)–(3) and (2)–(3) binaries are not easy to determine due to the large volatility difference of the components. Parameters should, therefore, be verified using data for the activity coefficients at infinite dilution of (1) in (3) and (2) in (3). These are relatively easy to measure and can in many cases be found in data banks and data compilations [51].

To illustrate the problem, Fig. 10 shows the separation factor between benzene (1) and cyclohexane (2) in the ternary mixture with the popular entrainer NMP (*N*-methyl-pyrrolidone) (3) at a constant pressure of 101.3 kPa. As can be seen in Fig. 10a (modified UNIFAC), adding NMP decreases the separation factor up to values below 0.4 at high NMP concentrations, thereby enabling simple separation of (1) from (2). This matches the findings from plant data.

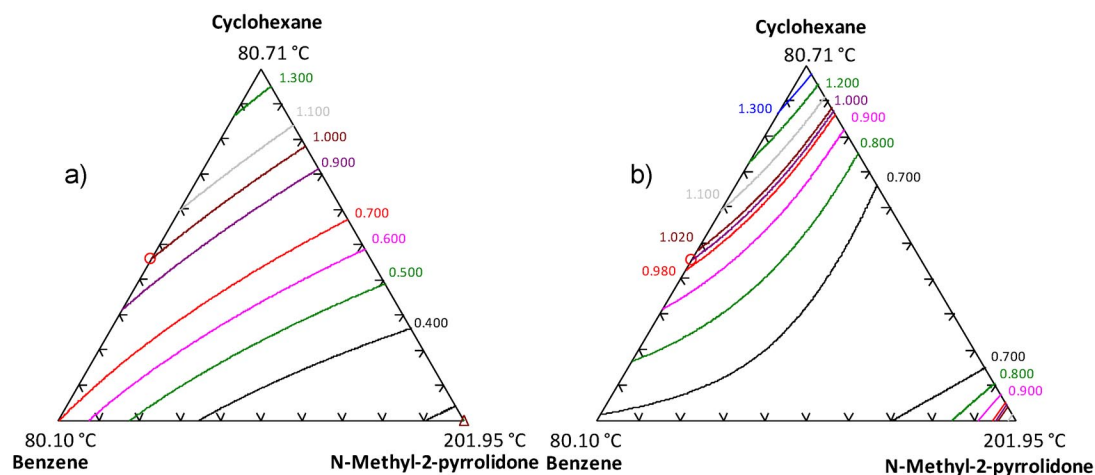


Fig. 10 Separation factor between benzene (1) and cyclohexane (2) in the ternary mixture with NMP (3) at a constant pressure of 101.3 kPa, (a) modified UNIFAC), (b) UNIQUAC with binary interaction parameters supplied by a process simulator.

Figure 10b shows the same calculations but at this time the UNIQUAC model with binary interaction parameters supplied by a process simulator was used. In case of b) the separation factor decreases to a value slightly below 0.7 and then increases again at high NMP concentrations while the binary azeotrope of the (1)–(2) mixture is predicted with nearly identical composition. Using these parameters would lead to very unrealistic simulation results.

When examining the results for the individual binary subsystems, it turned out that the binary UNIQUAC interaction parameters for the mixture cyclohexane with NMP are unrealistic at high NMP concentrations while the results of modified UNIFAC are satisfactory. As can be seen in Fig. 11, the limiting value of the separation factor between cyclohexane (1) and NMP (2) is only 17 instead of 40 in the case of modified UNIFAC. Correcting the parameters for this binary mixture leads to a correct description of the ternary data.

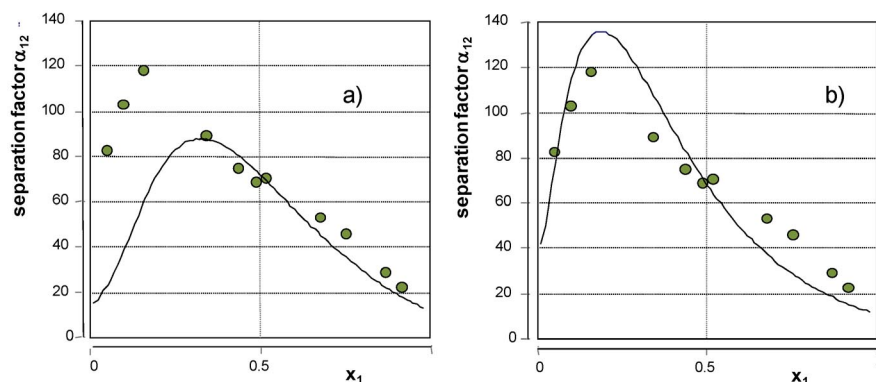


Fig. 11 Separation factor between cyclohexane (1) and NMP (2) at a constant pressure of 101.3 kPa (a) UNIQUAC with binary interaction parameters supplied by a process simulator, (b) modified UNIFAC, experimental data taken from the DDB [7].

USE OF FACTUAL DATA BANKS AND PREDICTIVE METHODS FOR PROCESS SYNTHESIS

Due to the different advantages of distillation, this separation process is also employed in the case of unfavorable separation factors. Different alternatives are available if ordinary distillation is not feasible:

- In some cases, the azeotropic behavior disappears at lower or higher temperature (pressure) and ordinary distillation is possible at these different conditions.
- In case the azeotropic composition varies strongly with temperature (pressure), the pressure swing distillation is often used. This situation arises mostly if the enthalpy of vaporization and consequently the slope of the vapor pressure curve of the two components to be separated differ significantly as for water plus organic components.
- In case the mixture forms a light boiling heterogeneous azeotrope, the distillate streams of two columns are condensed and the liquid phases are separated in a decanter and recycled to the columns. The pure compounds are obtained as the bottom products of the columns.

If neither of the above options solves the problem, separation can often be achieved by using an entrainer:

- In extractive distillation, a high boiling solvent is added slightly below the top of the column. Its task is to alter the ratio of the activity coefficients of the two components to be separated. In the case of the aliphatic-aromatic separation shown in Fig. 11, the interaction between NMP and the aromatic electrons of benzene lowers the volatility of benzene. At sufficiently high NMP concentrations, this results in a separation factor of less than 0.4 between benzene and cyclohexane. In a second column the selective solvent is recovered and recycled to the extractor column.
- In the case of azeotropic distillation, a suitable solvent is added which forms the lowest boiling azeotrope with one or both of the components to be separated. This light boiling azeotrope is obtained as the top product. In the case of a hetero-azeotrope, separation of entrainer and product is usually simple. In other cases, a similarly simple separation of this mixture has to be found.

A simple criterion to test the applicability of a component as extractive solvent is to calculate the activity coefficients at infinite dilution of the components to be separated in the entrainer. These should be as different as possible but high values of one of the activity coefficients (above 8 or 10) can lead to limited miscibility and therefore lower the capacity of the solvent.

Potential entrainers for azeotropic distillation can be found by examining available data on binary or ternary azeotropes or by estimating these data using predictive models.

In order to assist the engineer to find optimal entrainer candidates, a special program as part of the package “DDB Process Synthesis” has been developed [2]. The software first examines the behavior of the binary mixture that needs to be separated and verifies the existence of an azeotrope. It will then analyze if this separation problem can be solved without introducing a further component (at reduced or elevated pressure, hetero-azeotropic distillation or pressure-swing distillation). If this is not possible, the program proceeds by searching for a suitable entrainer by direct access to the DDB (>50 000 γ^∞ values, >50 000 azeotropic or zeotropic data) or results from predictive methods for a large number of potential solvents (up to 26 000). The detailed procedure is shown in Fig. 12 for both options.

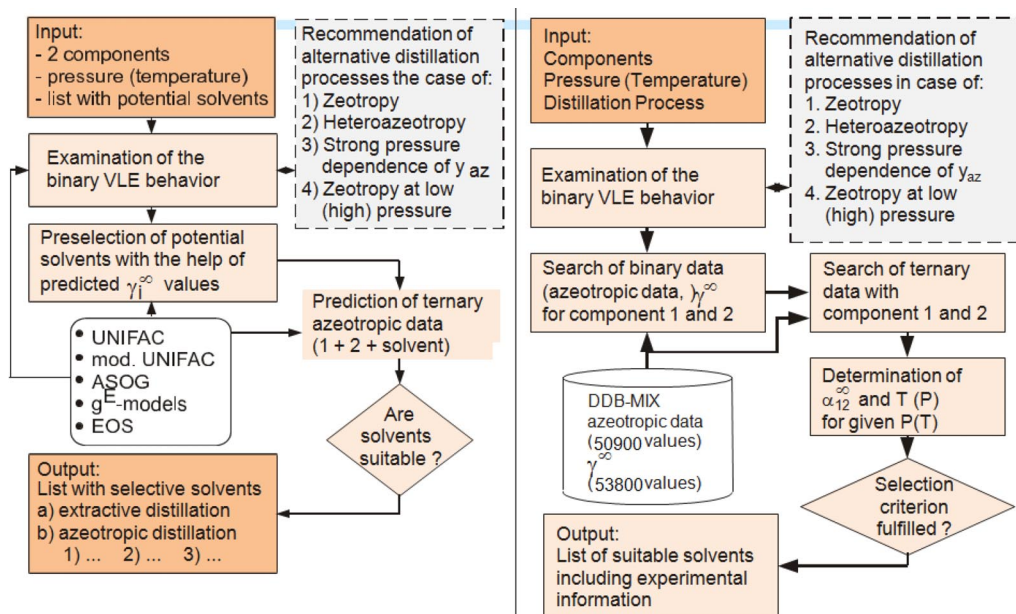


Fig. 12 Flowcharts of the two complementary methods for entrainer selection.

In the case of selection of suitable solvents by DDB access, information about the zeotropic/azeotropic behavior is taken from the database for azeotropic data (>50 900 data points). The database on activity coefficients at infinite dilution which contains 53 800 data points is used to account for information on selectivity and separability.

As this procedure can only detect entrainers for which data with the components to be separated are in the DDB, a second complementary approach was implemented, in which all 26 000 components can be used as potential suitable solvents. In this case, the required activity coefficients at infinite dilution (γ^∞) are estimated using models like UNIFAC, modified UNIFAC, and ASOG. Whether an azeotrope exists in a binary mixture can be decided from the pure-component vapor pressures P_1^S and P_2^S and estimated γ^∞ values of the two components* while calculation of the exact azeotropic concentration requires iteration.

*As discussed in the preface of the azeotropic data data collection [53], in case of azeotropic behavior of a binary mixture it must hold that either $\gamma_2^\infty > P_1^S/P_2^S > 1/\gamma_1^\infty$ or $\gamma_2^\infty < P_1^S/P_2^S < 1/\gamma_1^\infty$.

Both approaches are complementary in that the selection by DDB access is able to yield only components for which the required experimental information is available and the predictive approach can only cover components for which predictive models are applicable. It is therefore important to use both approaches to select the most suitable solvent. Some results of the entrainer selection for the systems ethanol–water and benzene–cyclohexane for both azeotropic and extractive distillation are shown in Fig. 13.

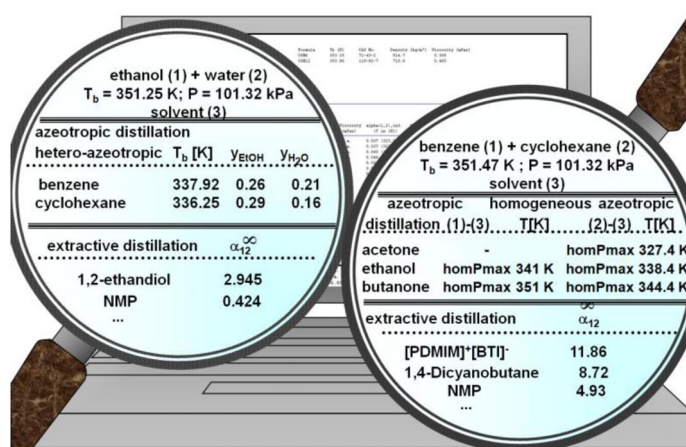


Fig. 13 Some results of the entrainer selection for the systems ethanol–water and benzene–cyclohexane for both azeotropic and extractive distillation by DDB access.

Figure 14 gives some results of the entrainer selection for the systems vinyl acetate–methanol for extractive distillation by DDB access.

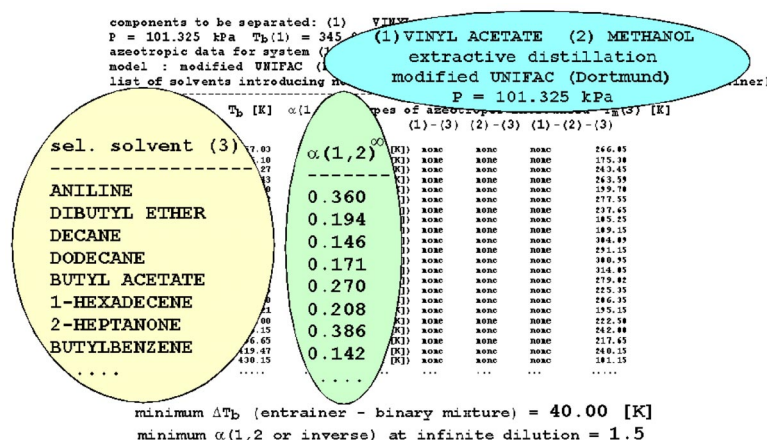


Fig. 14 Some results of the entrainer selection for the systems vinyl acetate–methanol for azeotropic and extractive distillation by DDB access.

Besides their effect on the mixture behavior, further properties of the entrainer are of technical importance (liquid viscosity, flammability, melting point, ...). Resulting component lists can optionally contain this information, and components with unfavorable properties can be removed.

FURTHER APPLICATIONS OF THERMOPHYSICAL DATA BANKS

In the past, industrial applications, especially process simulation and process synthesis, have served as the main driving force for the development of thermophysical data banks. In the case of the DDB where no public funding is available, the majority of revenues are generated from customers operating in this field. With thermophysical data banks and reliable estimation methods developed from the huge number of data readily available, other fields also start to benefit. Only a few applications should be mentioned here:

In environmental protection it is of great importance to be able to predict the fate of a chemical in the environment. A key property is the octanol–water partition coefficient, which is related to water solubility, soil/sediment adsorption coefficients, and bioconcentration factors for aquatic life. It allows judging the distribution of a component between the aqueous phase and, e.g., a fish. At the same time, it correlates with biotoxicity as components with very low and very high K_{OW} values have difficulties to cross the organic membrane of the cell. DDB and other thermophysical data banks contain numerous K_{OW} values, and several estimation methods are available [e.g., 38].

In order to calculate the distribution of a component between the aqueous phase and air, the vapor pressure of a component has to be known besides the activity coefficient at infinite dilution in water. Owing to the strong temperature dependence of the vapor pressure, the air/water distribution coefficient is significantly higher in the hot equatorial regions than at the very low temperatures close to the poles. Thus, volatile pollutants evaporate in warmer regions into the atmosphere and return to liquid and solid phases like lakes, soil, or animals in colder parts of the world. This leads to the observed accumulation of many pollutants in cold areas (“global distillation” [65]).

In atmospheric quality models, the amount of pollutants on aerosols is of great importance and estimation methods require the knowledge of the vapor pressure of a component and the activity coefficient at infinite dilution in the water film covering the aerosol particles [66]. Data of DDB and models developed on the basis of DDB data are intensively used for the development and test of such models.

In workplace safety, thermophysical property data and estimations are required to judge exposure to chemicals through inhaling [67,68] or diffusion through safety gloves [69]. Risk assessment models for chemical plants regularly employ PvT data, phase equilibria, and models developed with the help of DDB data (PSRK, VTPR) to estimate worst-case scenarios like pressure build-up in closed vessels.

Based on the equation of Le Chatelier (1891), flammability limits (flash points) of liquid mixtures can be calculated based on the pure-component flammability limits and the partial pressures of the components in the vapor phase above the mixture [70]. Calculation of the partial pressure above a liquid mixture requires knowledge about the VLE stored in thermodynamic data banks or available from modern predictive methods.

Life science companies nowadays show great interest in the selection of optimal solvents or solvent mixtures for pharmaceuticals and their intermediates. This has led to a first test of existing estimation methods based on data stored in the DDB [71].

Whenever new technologies or novel approaches become popular, the need for both data and predictive methods is obvious. Interest in ionic liquids as alternative “green” designer solvents led to measurement and publication of a large number of pure-component and mixture data for this chemical family. With the availability of sufficient data, work was started on predictive tools [72–74].

In many different fields, chemicals and their mixtures are used for specific purposes and components with special properties are required. Novel developments in refrigeration (e.g., using intermittent refrigeration cycles), reaction engineering where cosolvents enhance overall rates in biphasic reactions, ... require components with specific properties. For this purpose, computer-aided molecular design (CAMD) and data-mining methods are being developed that can generate or find molecular structures of components with the desired properties based on the knowledge coded in data banks and estimation methods.

The number of applications of thermophysical property data banks is rapidly growing, in part because modern data banks not only supply the available experimental results but also offer powerful estimation methods. Due to this predictive capability, they can supply more or less reliable information about new fields of scientific and technical development.

CONCLUSION

Factual data and data estimation methods are of great and increasing practical importance, whereby the development of reliable predictive methods would not be possible without the vast amount of experimental information stored in factual data banks. While the different data compilations developed up to about 1970 focused on supplying information to the user directly, modern electronic databases rely heavily on integrated software packages that feature complex data visualization, regression techniques, and export formats to other software products like process simulation software. In the future, an ever-stronger integration of data banks, integrated software, and applications can be expected. If experiences from other fields can be applied to sources for thermophysical properties, more and diverse forms of delivery and presentation (portals) will develop while the maintenance of the data collections and development of methods will remain the most cost- and labor-intensive task (content generation).

In the 21st century, simulation models are of predominant importance for scientific and technical developments and form the basis of economic and political decisions. Experiences from chemical process simulation show that inadequate basic data like thermophysical properties are the major reason for unrealistic results of simulation models, leading to customer frustration and erroneous business and political decisions.

Factual data banks and predictive methods combined with appropriate data-mining tools, therefore, carry a huge potential for economic and technological progress.

For a growing number of applications, thermophysical data banks and integrated reliable property estimation methods are of great importance and lead to improved processes with more efficient use of natural resources and less environmental hazard as well as rapid technological progress. New developments and technologies in many fields would not be possible without the vast amount of knowledge in these data compilations and methods, and it can be expected that the use of these tools both intensifies in core areas and expands to fields that still employ more empirical approaches to problem solution.

In the case of the DDB, the importance of further input and maintenance of the data bank, constant development and improvement of predictive methods for pure-component and mixture properties and application programs such as

- property estimation from molecular structure (ARTIST),
- verification of model parameters prior to process simulation, and
- process synthesis tools (Entrainer Selection)

were described in more detail.

Owing to the rapid development, the reader should use the links to the various Internet sites given in the text for the most up-to-date additional information.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Dr. U. Onken, Dr. H. Stage, Prof. Dr. Aa. Fredenslund, foreign guests, and numerous diligent Ph.D. students and co-workers for help in developing the DDB as well as the different thermodynamic models. Furthermore, we gratefully acknowledge the financial support received from Deutsche Forschungsgemeinschaft (DFG), Arbeitsgemeinschaft Industrieller Forschungsvereinigungen (AiF), Bundesministerium für Forschung und Technologie, South African National Research Foundation, Fachinformationszentrum Chemie, Berlin, and DECHEMA e.V., Frankfurt/Main.

REFERENCES

1. C.-C. Chen, S. Watanasiri, P. M. Mathias, V. V. de Leeuw. In *Chemical Thermodynamics in Industry*, T. Letcher (Ed.), pp. 166–179, Royal Society of Chemistry, Cambridge (2004).
2. J. Gmehling, C. Möllmann. *Ind. Eng. Chem. Res.* **37**, 3112 (1998).
3. J. Gmehling. *Fluid Phase Equilibr.* **144**, 37 (1998).
4. J. Gmehling. “Chemical Thermodynamics: The Basis for the Synthesis, Design and Optimization of Thermal Separation Processes”, in IUPAC “Chemistry for the 21st Century” monograph, T. M. Letcher (Ed.), pp. 1–13, Blackwell Science, Oxford (1999).
5. S. Horstmann, K. Fischer, J. Gmehling. *Chem. Eng. Commun.* **192**, 336 (2005).
6. U. Onken, J. Rarey-Nies, J. Gmehling. *Int. J. Thermophys.* **10**, 739 (1989).
7. J. Gmehling, J. Rarey, J. Menke. Dortmund Data Bank, Oldenburg (2009). <<http://www.ddbst.com>>.
8. D. A. Palmer. *Handbook of Applied Thermodynamics*, p. 41, CRC Press, Boca Raton (1987).
9. J. C. Wang, G. E. Henke. *Hydrocarbon Process.* **45**, 155 (1966).
10. L. M. Naphthali, D. P. Sandholm. *AIChE J.* **17**, 148 (1971).
11. G. M. Wilson. *J. Am. Chem. Soc.* **86**, 127 (1964).
12. H. Renon, J. M. Prausnitz. *AIChE J.* **14**, 135 (1968).
13. D. Abrams, J. M. Prausnitz. *AIChE J.* **21**, 116 (1975).
14. G. Soave. *Chem. Eng. Sci.* **27**, 1197 (1972).
15. D. Y. Peng, D. B. Robinson. *Ind. Eng. Chem. Fundam.* **15**, 59 (1976).
16. U. Westhaus, T. Dröge, R. Sass. *Fluid Phase Equilibr.* **158–160**, 429 (1999).
17. J. Barthel, H. Popp. *J. Chem. Inf. Comput. Sci.* **31**, 107 (1991).
18. Fachinformationszentrum CHEMIE, Berlin, Germany, <www.fiz-chemie.de>.
19. C. E. Garcia, D. M. Prett, M. Morari. *Automatica* **25**, 335 (1989).
20. J. H. Hildebrand, R. L. Scott. *Regular Solutions*, Prentice-Hall, Englewood Cliffs, NJ (1962).
21. E. L. Derr, C. H. Deal. I. *Chem. E Symp. Ser.* **32**, 3:40–3:51, Institute of Chemical Engineering, London (1969).
22. Aa. Fredenslund, R. L. Jones, J. M. Prausnitz. *AIChE J.* **21**, 1086 (1975).
23. Aa. Fredenslund, J. Gmehling, P. Rasmussen. *Vapor-liquid Equilibria Using UNIFAC: A Group Contribution Method*, Elsevier, Amsterdam (1977).
24. (a) H. Stage, P. Faldix. *Fortschr. Verfahrenstechnik* **2**, 306 (1956); (b) H. Stage, P. Faldix. *Fortschr. Verfahrenstechnik* **3**, 464 (1958); (c) H. Stage, P. Faldix. *Fortschr. Verfahrenstechnik* **4**, 429 (1961); (d) H. Stage, P. Faldix. *Fortschr. Verfahrenstechnik* **5**, 515 (1962).
25. J. Gmehling, U. Onken, W. Arlt, P. Grenzheuser, B. Kolbe, U. Weidlich, J. Rarey. *Vapor-Liquid Equilibrium Data Collection*, 31 parts, DECHEMA Chemistry Data Series, Frankfurt (1977–2008).
26. U. Weidlich, J. Gmehling. *Ind. Eng. Chem. Res.* **26**, 1372 (1987).
27. J. Gmehling, J. Li, M. Schiller. *Ind. Eng. Chem. Res.* **32**, 178 (1993).
28. Th. Holderbaum, J. Gmehling. *Fluid Phase Equilibr.* **70**, 251 (1991).
29. S. Horstmann, K. Fischer, J. Gmehling. *Fluid Phase Equilibr.* **167**, 173 (2000).
30. J. Ahlers, J. Gmehling. *Ind. Eng. Chem. Res.* **41**, 5890 (2002).
31. J. Li, H. M. Polka, J. Gmehling. *Fluid Phase Equilibr.* **94**, 89 (1994).
32. H. M. Polka, J. Li, J. Gmehling. *Fluid Phase Equilibr.* **94**, 115 (1994).
33. J. Li, Y. Lin, J. Gmehling. *Ind. Eng. Chem. Res.* **44**, 1602 (2005).
34. J. Huang, J. Li, J. Gmehling. *Fluid Phase Equilibr.* **275**, 8 (2009).
35. W. Yan, M. Toppoff, C. Rose, J. Gmehling. *Fluid Phase Equilibr.* **162**, 97 (1999).
36. H. K. Hansen, M. Schiller, Aa. Fredenslund, J. Gmehling, P. Rasmussen. *Ind. Eng. Chem. Res.* **30**, 2352 (1991).
37. A. Jakob, H. Grensemann, J. Lohmann, J. Gmehling. *Ind. Eng. Chem. Res.* **45**, 7924 (2006).

38. (a) G. Wienke, J. Gmehling. *Toxicol. Environ. Chem.* **65**, 57 (1998); (b) Erratum: G. Wienke, J. Gmehling. *Toxicol. Environ. Chem.* **67**, 275 (1998).
39. M. Kleiber. *Fluid Phase Equilib.* **107**, 161 (1995).
40. V. Liebert, S. Nebig, J. Gmehling. *Fluid Phase Equilib.* **268**, 14 (2008).
41. J. Li, M. Topp hoff, K. Fischer, J. Gmehling. *Ind. Eng. Chem. Res.* **40**, 3703 (2001).
42. H. Grensemann, J. Gmehling. *Ind. Eng. Chem. Res.* **44**, 1610 (2005).
43. T. Mu, J. Rarey, J. Gmehling. *Ind. Eng. Chem. Res.* **46**, 6612 (2007).
44. W. Cordes, J. Rarey. *Fluid Phase Equilib.* **201**, 397 (2002).
45. Y. Nannoolal, J. Rarey, D. Ramjugernath, W. Cordes. *Fluid Phase Equilib.* **226**, 45 (2004).
46. Y. Nannoolal, J. Rarey, D. Ramjugernath. *Fluid Phase Equilib.* **252**, 1 (2007).
47. Y. Nannoolal, J. Rarey, D. Ramjugernath. *Fluid Phase Equilib.* **269**, 117 (2008).
48. B. Moller, J. Rarey, D. Ramjugernath. *J. Mol. Liq.* **143**, 52 (2008).
49. Y. Nannoolal, J. Rarey, D. Ramjugernath. *Fluid Phase Equilib.* **281**, 97 (2009).
50. J. Gmehling, C. Christensen, P. Rasmussen, U. Weidlich, Th. Holderbaum. *Heats of Mixing Data Collection*, 4 parts, DECHEMA Chemistry Data Series, Frankfurt (since 1984).
51. J. Gmehling, J. Menke, D. Tiegs, A. Medina, M. Soares, J. Bastos, P. Alessi, I. Kikic, M. Schiller. *Activity Coefficients at Infinite Dilution*, 6 parts, DECHEMA Chemistry Data Series, Frankfurt (since 1986).
52. J. M. Sørensen, W. Arlt. *Liquid-Liquid Equilibrium Data Collection*, 4 parts, DECHEMA Chemistry Data Series, Frankfurt (1979).
53. J. Gmehling, J. Menke, J. Krafczyk, K. Fischer. *Azeotropic Data*, 3 parts, VCH, Weinheim (2004).
54. J. Gmehling. *J. Chem. Thermodyn.* **41**, 731 (2009).
55. K. G. Joback, R. C. Reid. *Chem. Eng. Commun.* **57**, 233 (1987).
56. W. Cordes, J. Rarey, F. Delique, J. Gmehling. *Software Development in Chemistry 7, Proceedings of Computer in der Chemie*, D. Ziessow (Ed.), Springer, Berlin (1993).
57. S. E. Stein, R. L. Brown. *J. Chem. Inf. Comput. Sci.* **34**, 581 (1994).
58. L. R. Constantinou, R. Gani. *AIChE J.* **40**, 1697 (1994).
59. J. Marrero-Morejon, E. Pardillo-Fontdevila. *AIChE J.* **45**, 615 (1999).
60. D. Ericksen, W. V. Wilding, J. L. Oscarson, R. L. Rowley. *J. Chem. Eng. Data* **47**, 1293 (2002).
61. J. R. Rowley, W. V. Wilding, J. L. Oscarson, R. L. Rowley. *Int. J. Thermophys.* **28**, 824 (2007).
62. I. U. Westhaus, R. Sass. *Fluid Phase Equilib.* **222–223**, 49 (2004).
63. W. V. Wilding, R. L. Rowley, J. L. Oscarson. *Fluid Phase Equilib.* **150–151**, 413 (1998).
64. M. Baerns, A. Behr, A. Brehm, J. Gmehling, H. Hofmann, U. Onken, A. Renken. *Technische Chemie*, Wiley-VCH, Weinheim (2006).
65. S. I. Sandler. *Fluid Phase Equilib.* **116**, 343 (1996).
66. S. L. Clegg, M. J. Kleeman, R. J. Griffin, J. H. Seinfeld. *Atmos. Chem. Phys. Discuss.* **7**, 11049 (2007).
67. J. Gmehling, U. Weidlich, E. Lehmann, N. Fröhlich. *Staub Reinhalt. Luft* **49**, 227 (1989).
68. J. Gmehling, U. Weidlich, E. Lehmann, N. Fröhlich. *Staub Reinhalt. Luft* **49**, 295 (1989).
69. A. Böning, J. Gmehling. *Durchlässigkeit von Schutzhandschuhen gegenüber Lösemitteln* Schriftenreihe der Bundesanstalt für Arbeitsschutz - Forschung - Fb 659, Dortmund (1992).
70. J. Gmehling, P. Rasmussen. *Ind. Eng. Chem. Fundam.* **21**, 186 (1982).
71. I. Hahnenkamp, G. Graubner, J. Gmehling. *Int. J. of Pharmaceutics*. Submitted for publication.
72. R. Kato, J. Gmehling. *J. Chem. Thermodyn.* **37**, 603 (2005).
73. S. Nebig, R. Boelts, J. Gmehling. *Fluid Phase Equilib.* **258**, 168 (2007).
74. S. Nebig, V. Liebert, J. Gmehling. *Fluid Phase Equilib.* **277**, 61 (2009).