

# From Raw Physical Data to Reliable Thermodynamic Model Parameters

Dr.-Ing. U. Westhaus<sup>\*</sup>, Dr. R. Sass

*DECHEMA Gesellschaft für Chemische Technik und Biotechnologie e.V.,  
Theodor-Heuss-Allee 25, D-60486 Frankfurt am Main, Germany*

## Abstract

Nowadays actual process simulation packages are reliable and valuable tools for many tasks occurring in the day of a chemical process engineer. However, the accuracy of a process simulation depends strongly on the thermodynamic models used to describe the physical behavior of the involved components. To achieve real world behavior with a model one therefore needs both reliable property data as well as a full featured data regression package.

One example for a valuable source of data is the numerical database DETHERM. It contains thermophysical property data for the design and optimization of chemical plants. At the time more than 4.2 million data sets for more than 122,000 pure components and mixtures are stored. The properties stored include mainly phase equilibrium data (VLE, LLE, VLLE, GLE and SLE), vapor pressures, critical data, thermodynamic properties, transport properties, surface tensions and electrolyte data.

Following up to database retrieval the DECHEMA Data Preparation Package closes the gap between raw thermophysical data and model parameters, as used in any process simulation package. Retrieved or otherwise generated data sets can be displayed, compared and selected. Afterwards parameters for pure component as well as binary models can be regressed and displayed in comparison with the raw data. The regression module is capable to correlate simultaneously different data types using different error functions. For example the simultaneous correlation of VLE, LLE, HE,  $\gamma$ ,  $\gamma$ -infinity and azeotropic data is possible. The package is not restricted to its built-in thermodynamic models. Thermodynamic calculations are performed using a standardized interface. This enables users to simply plug in any thermodynamic package and/or model of their own choice.

## 1. Introduction

Process synthesis, design and optimisation and also detail engineering for chemical apparatus depend heavily on availability and reliability of thermophysical property data for the pure substances and mixtures involved. If data are not available, no process simulation or feasibility study is possible. If the data are not reliable, proper operation of apparatus cannot be guaranteed and costs figured out in advance may just be worthless numbers. Without access to a numerical database and if the available literature and notes do not contain a value, the only possibilities are to measure the data or to calculate

---

<sup>\*</sup> Corresponding author: Fax: ++49-69-7564-201; e-mail: westhaus@dechema.de

them with a group contribution method or another estimation routine. The first alternative will be expensive and time consuming, the second one will produce data with unknown reliability in most cases. Therefore access to a thermophysical property database will be first choice. DETHERM<sup>®</sup> is such a numerical database package containing thermophysical properties of pure substances and mixtures. DETHERM<sup>®</sup> is an acronym for DECHEMA THERMOPHYSICAL PROPERTY DATABASE. It is one of the worlds largest databases in this field. Today 4.2 million data sets for around 20,000 pure components and 100,000 mixtures are stored in DETHERM<sup>®</sup>. The database is updated yearly. Parts of the data are also published in printed form as the "DECHEMA Chemistry Data Series".

## 2. DETHERM<sup>®</sup> Database contents

The DETHERM<sup>®</sup> database is jointly produced from the DDBST GmbH in Oldenburg, the FIZ CHEMIE in Berlin, the University of Regensburg and the DECHEMA e.V.. The database consists of several property orientated packages. Each of these packages is produced and maintained from groups having expertise in the respective property area since years. Performing consistency and error checking is self-evident upon data entry in this groups. The main property areas DETHERM<sup>®</sup> contains data for, are:

- phase equilibrium data (VLE, LLE, VLLE, GLE, SLE)
- vapor pressures, critical data
- thermodynamic properties
- transport properties
- surface tensions
- electrolyte data

Each of the data records contains the numerical data itself, a bibliographic reference, an abstract and usually some index terms.

## 3. DETHERM<sup>®</sup> access options

To access the DETHERM<sup>®</sup> database two different options are available. If data are needed only sporadically, "**DETHEM ... on the WEB**" is the most suitable solution. For more frequent use, the in-house system **DETHEM-ORS** is the first choice.

### 3.1 "DETHEM ... on the WEB"

"**DETHEM ... on the WEB**" is DECHEMA's web-based solution to access the DETHERM<sup>®</sup> database<sup>1</sup>. This service is designed to access quickly all available data of the searched substances or mixtures. Searches may be restricted to special properties (e.g. vapor-liquid-equilibrium data) and/or to limiting temperature and pressure ranges.

---

<sup>1</sup> Its URL is "<http://i-systems.dechema.de/detherm>".

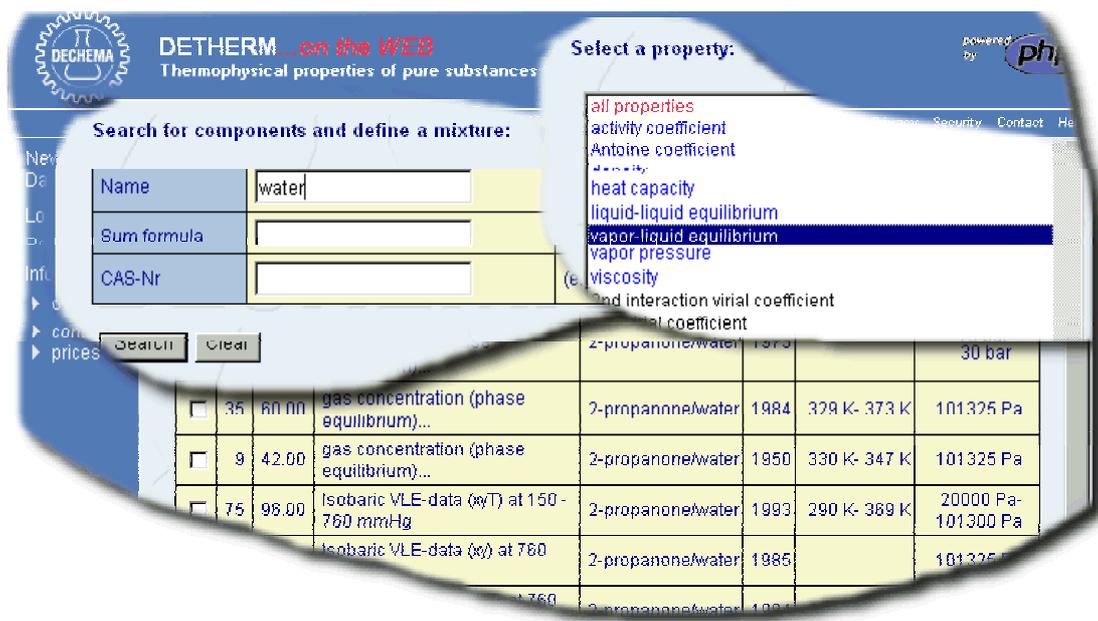


Fig. 1. DETHERM ... on the WEB

The result of a search request will be a list of data sets containing the following information (see also Fig. 1):

- number of lines in the data table
- costs of the data table in case of download
- description of the contents of the data table (heading of table)
- related mixture or pure substance the table contains data for
- the publication year of the literature reference
- the respective temperature and pressure ranges of the data

The above mentioned information and also searching the database as described is free of charge and available to everyone. To download data sets an account is required. In that case one gets:

- the numerical data of the data table
- the complete literature reference
- the complete substance identification

The data may be downloaded in a variety of different formats optimized for either direct viewing (HTML) or for direct usage in follow-up applications like process simulation packages (e.g. ASPEN Plus) or data regression tools (e.g. the DECHEMA Data Preparation Package).

### 3.2 The in-house solution DETHERM-ORS

In comparison to “**DETERM ... on the WEB**” as described above, the in-house solution for the DETHERM<sup>®</sup> database is designed for frequent access and usage in

multi-user environments. It consists of an ORACLE database server (containing the data) in combination with a specialized retrieval software (DETERM-ORS) running on each users client. The communication between clients and server is handled via network. In comparison to the web-based solution, the in-house systems offers the following main advantages:

- Easy-to-use interface to carry out the formulation of retrieval questions
- Direct graphical representation of data
- Possibility of data maintenance
- Possibility to perform “inventor” questions

The opportunity of direct graphical representation gives users an overview of the source distribution of the data sets (see Fig. 2). This makes viewing of general property behavior (e. g. dependency on temperature and/or composition) an easy task.

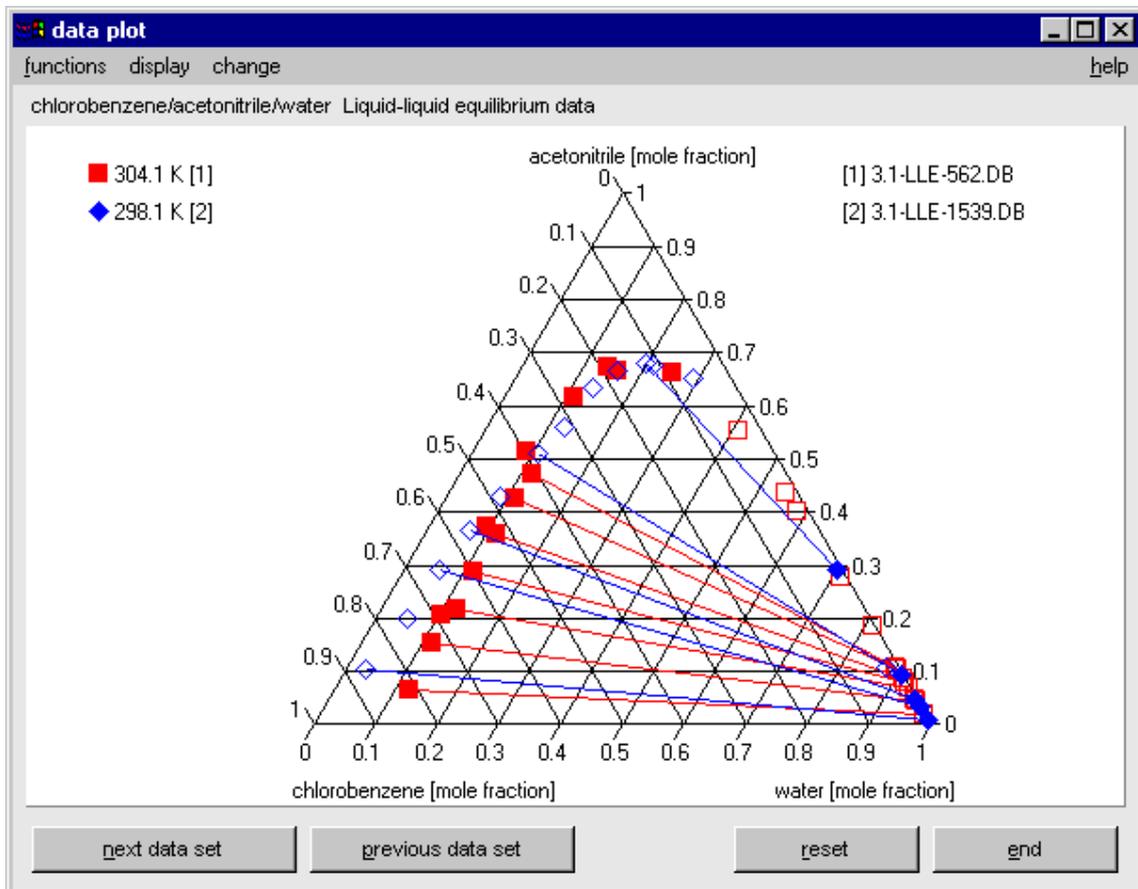


Fig. 2. Example plot from DETERM in-house system

The in-house system offers the possibility of data maintenance. This allows users to append own data to the database and enables companies to merge their own data pool with the one delivered. These data can be retrieved consistently together with the rest of the database, as mentioned before. The maintenance of literature and components or mixtures is also feasible.

Furthermore it is possible to run inventor questions, that is to search the database for substances or mixtures with given properties. Typical examples are searches for liquid substances having a thermal conductivity bigger than X cP in a given temperature range and a melting point bigger than Y °C or searches for substances building up a mixture gap with the miscible mixture A/B and also forming an azeotrope with component A. Running inventor questions against large databases offers significantly reduced response times in comparison to guessing by human experts.

## 4. From thermophysical data to thermodynamic model parameters

### 4.1 Introduction

Reliable thermophysical property data are only the first step to successful modelling of complete chemical processes. Model parameter regression typically follows up to data retrieval, because process simulation packages are not able to deal directly with discrete data points like measurement values. The step of choosing a property model and the related parameter regression is crucial for the overall quality of a process simulation. Because if the underlying thermodynamic models do not accurately reproduce the true physical behaviour of the involved components, the complete simulation will generate useless numbers only.

The DECHEMA **Data Preparation Package** closes the gap between raw thermophysical data (such as contained in the DETHERM<sup>®</sup> database) and model based process simulation packages (such as e.g. Aspen Plus<sup>®</sup>). It allows:

- Selection, graphical display and analysis of data sets
- Regression of model parameters
- Comparison of different models among each other

The Data Preparation Package is build strictly modular and consists of five main components (see Fig. 3). Following a typical project course, the components will be used in the manner described in the following.

The very first step in a project will be to load raw thermophysical property data, one wants to generate models for. With the **Neutral File Interface** one can to read and write raw thermophysical data and model parameters using the neutral IK-CAPE PPDX format. Data from a variety of databases can be assembled therefore. Regressed model parameters could afterwards be exported to a file and loaded from the targeted process simulation package.

The second step will be data visualization. The **Graphics Subsystem** allows the graphical representation of measured raw data and also model derived data in variety of different plots. Among them are:

- **Binary VLE:** TPXY, TXY, PXY, K-Value vs. X, McCabe-Thiele, Redlich-Kister, VanNess-Smith-Abbott
- **Binary LLE:** TX

- **Ternary LLE:** Orthogonal & equilateral triangles, distribution ratios

Outliers and (in some cases) the thermodynamic consistency of the different data sets and points could easily be identified with the help of such plots. Directly from the plot the weights of individual data points as well as the weights of completed data sets could be accessed and altered. Setting of accurate weights is a significant prerequisite in preparation of a data regression. It allows to eliminate outliers or to mark critical regions like e.g. azeotropic compositions with increased weights.

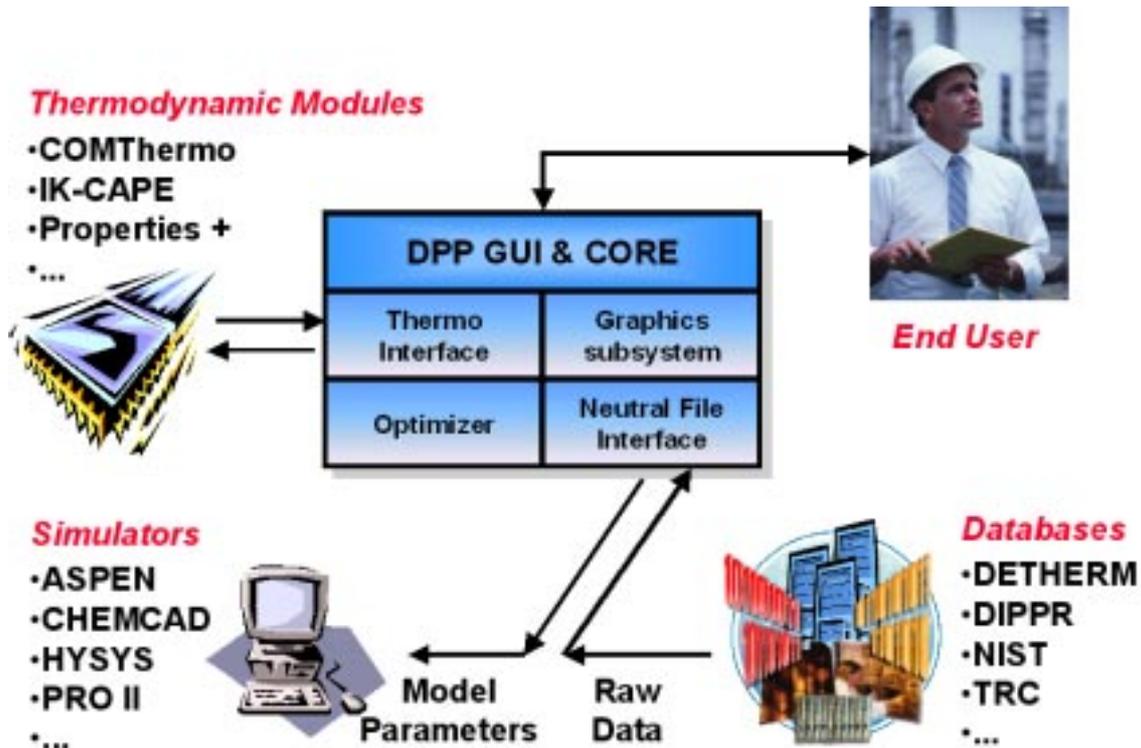


Fig. 3. General architecture of DECHEMA Data Preparation Package

In the third step one has to choose the thermodynamics package and the models to work with. The Data Preparation Package does not contain a thermodynamics module inside. Instead any thermodynamics module having an open interface can be used together with the package. Currently the IK-CAPE as well as the Aspen Properties<sup>®</sup> thermodynamics packages are fully supported. Interfaces to other packages are in preparation. When setting up models for parameter regression, all knowledge about the models (number of parameters, type of parameter, boundaries, meaningful starting values, etc.) must be extracted from the thermodynamics package utilizing the **Thermo Interface**.

The last step will be model parameter regression. For regression the internal **Optimizer** of the Data Preparation Package is used. During optimization the property calculation is performed with help of the external thermodynamics package, using the thermo interface. With results presents, the overall quality as well as a comparison of different models could be examined either by plots or textual reports.

## 4.2 Thermodynamic consistency by means of extended parameter regression

For the reliable description of a mixtures real behavior over the whole concentration and also a large temperature range, model parameters should be derived by simultaneous correlation of all available data. If, for example, during the design of a distillation tower only the VLE data are taken into account, a simulation will most probably model the required number of physical stages correctly. But the energy balance around the distillation tower may contain a systematic error, because the heats of mixing for the mixture may be predicted totally wrong. Therefore the usage of thermodynamically consistent models describing correctly all known properties of a component or mixture, are strongly recommended. Activity coefficient models, which are typically used for calculation of vapor-liquid- and liquid-liquid-equilibrium, heats of mixing, azeotropic points and  $\gamma^\infty$  data, are the most important class here.

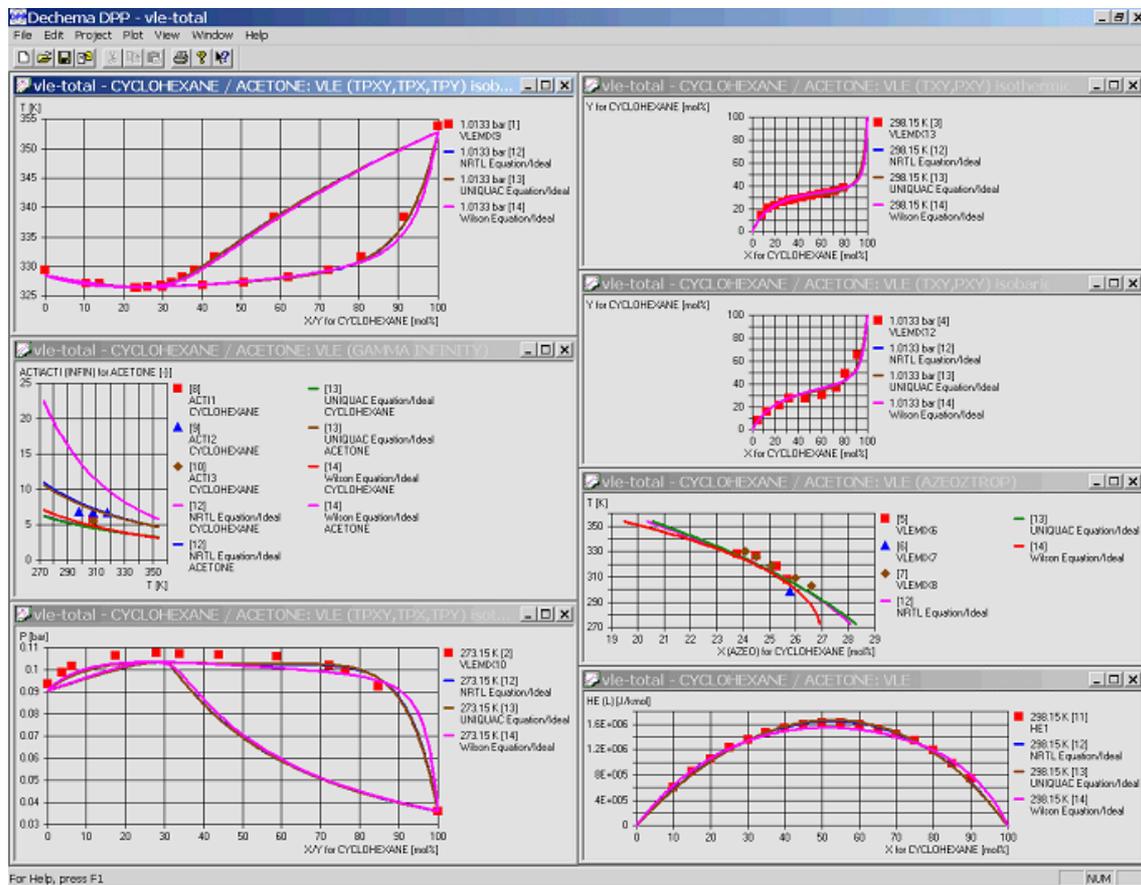


Fig. 4. Simultaneous regression of different of VLE data types,  $\gamma^\infty$ -data, azeotropic data and heats of mixing for the system cyclohexane/acetone at 25 °C. Compared are the results for the UNIQUAC, Wilson and NRTL-model

To enable users to perform simultaneous correlations of different data types, the optimizer of the DECHEMA Data Preparation Package is able to calculate a common objective value over  $n$  sets of data grouped in  $m$  blocks. For each of the  $m$  blocks a separate objective and error function could be used. To minimize the calculated global objective value, the model parameters are varied. In addition this methodology could be

combined with a Maximum-Likelihood approach for generation of the “true” values from the erroneous measurement values. This approach enables the Data Preparation Package to correlate different data types using different error functions simultaneously and allows to correlate VLE, LLE, HE,  $\gamma^\infty$  and azeotropic data of a mixture all together (see Fig. 4).

#### 4.3 Combination of different optimization approaches

One of the most challenging tasks during an optimization is to achieve convergence in a reasonable time. The DECHEMA Data Preparation Package included five different optimization algorithms, that could be started either mutually exclusive or in successive combinations. The most promising combination is to use at first a stable, but slow algorithm (like Simplex-Nelder-Mead or Powell algorithm) for generation of the first meaningful values from the models initial starting parameters. After a certain number of steps the algorithm could then be replaced with a faster one (like Gauß-Newton or Broyden-Fletcher-Goldfarb-Shanno algorithm), that typically requires more reasonable starting values.

If an optimization tends to lead to illegal parameters because of boundary violations, the parameters will be automatically transformed by the package using an periodic transformation function. Such a transformation alters an bounded parameter space to an unbounded one and makes boundary checking during optimization superfluous.

### 5. Conclusion

The DETHERM<sup>®</sup> database offers access to wide variety of thermophysical property data. One of the main focuses of DETHERM<sup>®</sup> are phase equilibrium data. The various access options are tailored to meet the needs of a wide range of different users. Interfaces to data regression tools and process simulation packages offer especially chemical engineers an integrated environment for data retrieval, maintenance and preparation.

The DECHEMA Data Preparation Package is typically used subsequent to data retrieval and closes the gap between raw thermophysical data and thermodynamic model parameters as needed for any chemical process simulation. Besides selection und graphical display of data sets the package permits the regression of model parameters as well as the comparison of models with each other. Simultaneous correlation of different data types allows generation of thermodynamically consistent model parameters for the description of all related properties. During optimization the thermodynamic calculations are performed using a standardized interface to allow usage of any thermodynamic package and/or model.