

# Digitalization in Thermodynamics

Esther Forte<sup>1</sup>, Michael Bortz<sup>2</sup>, Jakob Burger<sup>3</sup>, Jadran Vrabec<sup>4</sup>, and Hans Hasse<sup>1\*</sup>

<sup>1</sup>Laboratory of Engineering Thermodynamics (LTD), University of Kaiserslautern, Erwin-Schrödinger-Str. 44, 67663 Kaiserslautern, Germany

<sup>2</sup>Fraunhofer Institute for Industrial Mathematics (ITWM), Fraunhofer-Platz 1, 67663, Kaiserslautern, Germany

<sup>3</sup>Chair for Chemical Process Engineering, Technical University of Munich, Schulgasse 16, 94315, Straubing, Germany

<sup>4</sup>Thermodynamics and Energy Technology (ThEt), University of Paderborn, Warburger Str. 100, 33098, Paderborn, Germany

\*Correspondence: hans.hasse@mv.uni-kl.de.

## Abstract

Digitalization is about data and how they are used. This has always been a key topic in applied thermodynamics. In the present work, the influence of the current wave of digitalization on thermodynamics is analyzed. Thermodynamic modeling and simulation is changing as large amounts of data of different nature and quality become easily available. The power and complexity of thermodynamic models and simulation techniques is rapidly increasing, and new routes become viable to link them to the data. Machine learning opens new perspectives, when it is suitably combined with classical thermodynamic theory. Illustrated by examples, different aspects of digitalization in thermodynamics are discussed: strengths and weaknesses, as well as opportunities and threats.

**Keywords:** digitalization, thermodynamic models, Pareto optimization, uncertainty propagation, machine learning.

## 1. Introduction

Digitalization has a rapidly increasing impact on our economies and societies [1], [2]. It started with the advent of digital computers in the 1950s and has had an important influence in many fields since then. But it was not until recently that a comprehensive connectivity between fields like process design, process control, accounting, and marketing, has become feasible and is being developed at a high pace. The consequences are disruptive and this phase of digitalization has therefore been labeled as digital revolution, fourth industrial revolution, or industry 4.0. Digitalization potentially merges physical, biological, and social realms with those of mathematics and computer science, impacting all disciplines [3]. Paving the way for such a digital transition still requires much effort. Neither the route nor the outcome is fixed, and they will be shaped by those who engage in the endeavor.

Digitalization has had an important impact on chemical engineering and thermodynamics since computers became broadly available in that field in the 1960s. The impact the introduction of digitalization had can be described as disruptive: design of equipment could now be done computer-based, multicomponent mixtures could be simulated, relying on improved thermodynamic models, and powerful group contribution methods could be developed.

The role of computers in chemical engineering and thermodynamics has always been acknowledged and appreciated. In 1977, the Elsevier Journal “Computers and Chemical Engineering” was launched. In the early 1980s books were published with titles like: “Computer calculations of multicomponent vapor-liquid and liquid-liquid equilibria” [4]. Especially molecular thermodynamics was revolutionized by computer simulations, the first ones being carried out already in the 1950s [5]. Neural networks have been used in thermodynamics since the 1990s [6]. Moreover, machine learning is closely related to adjusting model parameters, which is a core business of thermodynamicists.

So, one can ask: what is new in the current digitalization hype? Our answer is the same for digitalization in thermodynamics as it is for digitalization in general: it is not so much the progress in the individual domains but the fast-paced merge between them. This creates a special momentum which might well turn out to be disruptive, also in thermodynamics.

In the present paper, we discuss different topics of digitalization in thermodynamics, focusing on new developments. Examples are given that stand for more general points, hoping that their sum offers a broad view of the field, without, however, striving for completeness. The paper is organized as follows: in Section 2, the stage is prepared by some remarks on reality, classical experiments, models, and computer experiments. In Section 3, new routes in thermodynamic modeling are discussed and new challenges are highlighted in Section 4. Experimental data are the topic of Section 5. Section 6 addresses the problem of dealing with those parts of a problem on which only insufficient knowledge is available. Finally, we take a preliminary look into machine learning in thermodynamics in Section 7, before conclusions are drawn in Section 8.

## **2. Reality and classical experiments, models and computer experiments**

An overview of computer-based model development is given in Figure 1. It is general and holds in particular also for the development of thermodynamic models. Let the goal be to develop a mathematical model for some property of a real substance  $x^{\text{real}}$ . The corresponding property described by the model is designated as  $x^{\text{mod}}$  here. As a result of digitalization, current models are generally so complex that computer simulations are needed for their evaluation. Algorithms have to be selected

and code has to be written and implemented on some computer before the simulation can be carried out. The result of the computer simulation is  $x^{\text{sim}}$ . The number for  $x^{\text{sim}}$  can be compared to a corresponding experimental result  $x^{\text{exp}}$ , and a feedback loop can be closed in which the model is adapted such that the deviations between  $x^{\text{exp}}$  and  $x^{\text{sim}}$  are minimized. This can be done either by structural changes in the model or by suitable parameterization.

Classical experiments and computer simulations have much in common, which is why the latter are often called computer experiments. In the classical experiment the reality is studied, in the computer simulation a model of the reality. Both the classical experiment and the computer simulation are prone to errors. In general, only the results of the experiment  $x^{\text{exp}}$  are known, but not the true value of the corresponding real-world property  $x^{\text{real}}$ . The same holds for the simulation: in general, only the simulation result  $x^{\text{sim}}$  is known, but not the true value of the property of the model  $x^{\text{mod}}$ .

The schematic shown in Figure 1 represents the situation that has developed as a result of the first wave of digitalization. It is now changing: the borders between the blocks begin to fade.

((Figure 1))

### **3. New routes in thermodynamic modeling**

#### **3.1. Multi-criteria optimization**

As a result of digitalization, paradigms in model development change. Routes become feasible that were inaccessible in the past. Assume that the task is to develop a thermodynamic model, e.g. an equation of state for a given real fluid. Traditionally, the outcome of the model development feedback-loop shown in Figure 1 is a certain set of parameters, which is found from an optimization with a single objective function. Often different data sets of different nature are used for the fit, which are lumped together in the objective function. However, the different data sets can generally not be represented equally well with the model. Hence, there are conflicting objectives. Such optimization tasks should instead be solved by multi-criteria optimization (MCO).

MCO permits finding sets of best compromises for a problem with conflicting objectives. The solutions to the MCO problem are known as Pareto-optimal solutions and represent points in the objective space for which an improvement of any of the objectives is not possible without deteriorating at least one other objective. The set of Pareto-optimal solutions (the so-called Pareto set) is often also referred to as Pareto front of the MCO problem. Efficient adaptive algorithms are available for finding the Pareto-set of a given problem [7].

A number of recent studies demonstrate the applicability and advantages of MCO for the development of thermodynamic models. They cover the most important types of thermodynamic models, namely

equations of state [8], models of the excess Gibbs energy of mixtures ( $G^E$ -models) [9], and molecular models (force fields) [10]–[12]. The knowledge of the Pareto front gives a comprehensive overview of what can be achieved with a given model regarding the description of the considered data sets. This is a big advantage compared to having only the information on a single optimal point supplied by the traditional single-objective approach. The MCO approach is general and has also been used in product and process design [13]–[18] and in optimal design of experiments [19].

The shape of the Pareto front provides the decision-maker with an overview of the trade-offs between the different objectives. In a two-dimensional problem, the trade-off is the deterioration that needs to be accepted in the value of one objective for a given improvement in the other one. An example is shown in Figure 2. The optimization problem studied in Figure 2 is the development of PC(P)-SAFT [20]–[22] models for water, details are reported in [8]. The conflicting objectives are the average deviation between model and experiments in the vapor pressure and the saturated liquid density, respectively. The trade-off between these two objectives is quantified for different PC(P)-SAFT versions. Particularly attractive solutions, if interest is in the simultaneous description of both objectives, are found in the so-called “Pareto knee”. It is sharp for two of the three model versions, cf. Figure 2. The model versions differ in the way the polarity is accounted for. The first model is the non-polar original PC-SAFT equation of state [20], [21], in the two other models, the polarity is taken into account by using PCP-SAFT [22]. They differ in the way, the magnitude of the dipole moment was determined. In one case, it is simply adopted from the literature [23], in the other it is used as an additional adjustable parameter. In Figure 2 it is shown in a comprehensive manner how well the experimental data can be described by these three variants of the model. In the traditional approach, usually only three points in that diagram would have been available for a comparison, representing three “best” individual parameterizations of each model. From Figure 1 it is clear that the polar version with fixed dipole moment generally gives an improvement over the non-polar version, even though the number of adjustable parameters is the same. The exception is, however, one point in the Pareto knee, where both types of model yield similar results. As expected, the polar version with the adjusted dipole moment leads to an improved description. It is also interesting to compare the pictures of the Pareto fronts in the parameters space. For the sake of brevity, the reader is referred to [8] for this discussion.

In a final step, the decision-maker must select a model from the Pareto set, which is most suited for his application. Non-trivial questions are connected to reporting the results from such MCO optimizations: while in the traditional approach, only a single parameter set had to be specified, and detailed results for it could be reported, in the MCO approach, the result is the Pareto front, which includes in principle an infinite number of models. Furthermore, the Pareto front has to be distinguished from its numerical approximations that are supplied by the MCO algorithms.

((Figure 2))

### 3.2. Combining results from classical and computer experiments to hybrid data sets

The development and parameterization of high-quality equations of state for describing real fluids requires a large amount of thermodynamic data. Usually, such data are exclusively generated by experimental studies, which are time-consuming and expensive. Furthermore, the range in which experimental data can be sampled may be limited by decomposition, hazards, etc. These difficulties are reflected in the limited number of substances for which high-quality equations of state are available. Molecular modeling and simulation has evolved in the last decades into a viable alternative to generate thermodynamic data of fluids. Molecular models are known to have powerful predictive capabilities, and simulation data can generally be generated much easier than experimental data. The quality of the simulation data depends, however, on the quality of the underlying molecular force field. It has recently been shown that it is attractive to combine molecular simulation data and experimental data for the development of high-quality equations of state: Equations of state for ethylene oxide [24], hexamethyldisiloxane [25], octamethylcyclotetrasiloxane [26] and dichloroethane [27] have been developed in this way. In all cases, scarce experimental data were supplemented by molecular simulation data to yield a hybrid data base.

An efficient method to generate independent thermodynamic data for this purpose is given by the statistical-mechanical formalism proposed by Lustig [28], [29] that has been integrated into the molecular simulation tool *ms2* [30], [31]. With this formalism, any derivative of the Helmholtz energy can be obtained by a single canonical ensemble simulation at a given state point.

Developing an equation of state for Ethylene oxide is an excellent example to illustrate the benefits of using a hybrid data base. Although it is an important intermediate in the production of many chemicals, only few experimental thermodynamic data are available in the literature because of its hazardous nature. A very accurate molecular force field, which won the fourth industrial fluid property challenge [32], is available for this substance [33]. In this case, molecular simulation data are mainly used to supplement experimental data in the homogeneous liquid and supercritical region. As a result, the range of validity of the new hybrid equation of state is doubled in terms of the absolute temperature and increased seventyfold in terms of pressure [24].

Another example for the applications of hybrid data sets is the development of an equation of state for hexamethyldisiloxane (HMDS), a working fluid for organic Rankine cycles that is described in [25]. An equation of state for HMDS by Colonna et al. [34] showed significant deviations from new experimental data for the speed of sound, as well as from molecular simulation data. The molecular model of HDMS was developed based on a limited amount of experimental vapor-liquid equilibrium

data. The deficiencies of the equation of state of Colonna [34] were attributed to the fact that the experimental data base for HMDS is too narrow. To extend that data base, systematic molecular simulations were performed and used together with the existing thermodynamic data for developing an improved equation of state. To illustrate the results, data for the speed of sound of HMDS that were obtained from experiments, molecular simulations, and from the new equation of state are shown in Figure 3. For comparison, also results from the equation of state of Colonna are displayed. The molecular simulation data agree with the experimental data within their statistical uncertainty, and the new equation of state [25] describes these data well, whereas the predictions with the equation of state of Colonna et al. [34] deviate significantly from the other data sets for all temperatures except for 573 K.

((Figure 3))

It could even be considered to use only simulated data for the development of an equation of state. This could be particularly interesting when no or hardly any experiments are available, the time-to-solution is critical, and a simple equation of state is sufficient.

Developing equations of state was considered to be a tedious task. There are many PhD theses in which only one equation of state for a single fluid was developed. Digitalization enables automation of many parts of this process. This has been examined in a recent study on phosgene [35]. In a first step, every task in the creation of the equation of state that requires considerable manual effort or expertise was identified. Such tasks were then simplified and automatized, wherever this was possible. As a result, a cloud-based workflow with a graphical user interface was developed which reduces the complex task of fitting equations of state to a few clicks [36].

#### **4. Challenges in thermodynamic modeling and simulation**

##### **4.1 Proliferation of model variants and epistemic value of models**

Increasing computing power and improved algorithms make more complex thermodynamic models feasible. On the other hand, the easiness to develop and modify such models has led to a plethora of models that serve similar purposes and differ only in details. Some of them are solely used by the group by which they were developed. This is obviously a problem as there is no independent testing. It would be desirable to embank this flood. Concentrating the efforts of the thermodynamic community on a few particularly promising thermodynamic models and pushing them forward together would be highly attractive both from a practical and a scientific standpoint.

The easiness with which model parameters can be adjusted has also encouraged over-parameterization and moving from physical models to correlations. Here, we define a “model” as carrying epistemic

value (i.e. giving insight), whereas a “correlation” does not carry such value (apart from the fact that it can be plugged into the framework of thermodynamic theory). In general, digitalization may have different effects on the borderline between models and correlations. Digitalization can shift the scales in the direction of increasing epistemic value, as more sophisticated models may become feasible. But the easiness of parameterization can also produce the reverse effect. Machine learning in its pure form is an extreme example of correlating data, as discussed in more detail in Section 7.

## 4.2 Simulation errors

Molecular simulation data are often considered to be “exact” [37], [38], as opposed to results from theories which are always based on some approximations in their derivations. Here, the term “theory” refers e.g. to theories for some model substance, e.g. the hard-body fluid or classes of processes like nano-scale flow. The approximations on which such theories are based can be checked by comparison with computer simulation data.

The existence of statistical errors is accepted for computer simulations and they are regularly quantified and reported. On the contrary, systematic errors, which are a well-known second type of uncertainty in classical experiments [39], are not generally accepted in computer simulation and sometimes considered to be simply a result of avoidable faults [40]. Systematic errors are hard to assess. A well-known strategy from the field of experimental work is to carry out round-robin studies in which different laboratories solve the same task with different equipment. This approach can also be used for computer experiments.

In a recent round-robin study [41], a comparison of molecular simulation results that were obtained by five research groups which worked independently is presented. Eight different well-established molecular simulation codes were used. All groups were given identical simulation tasks. They consisted in determining the density and the potential energy of four simple alkanes on a given temperature-pressure grid. Three different types of molecular models with internal degrees of freedom were considered.

The deviations between the results from the different groups were found to exceed the statistical uncertainty of the individual results, in many cases to a large extent. This even holds for the case when the same simulation code was used by different groups. An example is shown in Figure 4 for the determination of the density of *i*-butane at 41 MPa using the OPLS force field. The results obtained are depicted in terms of their relative deviation from their arithmetic mean. The arithmetic mean is interpreted here as a guess for the true value. The deviations from that mean are as large as 0.8 % for the lowest temperature. Also in most other cases, they exceed the statistical uncertainty of the individual results, which is often below 0.05%.

Undoubtedly, avoiding systematic errors in simulations should always be the goal. The study emphasizes the challenges that have to be met in trying to achieve that goal. Much can be learnt here from experimentalists, e.g. carrying out tests of the equipment that is used by comparisons with accepted benchmarks. The present study also supplies such benchmark data.

((Figure 4))

## **5. New types of experimental data**

### **5.1 Ubiquitously available data**

One of the most popular narratives in discussion on digitalization is about ubiquitously available data, i.e. masses of data which only wait to be analyzed to turn the knowledge into gold. Let us briefly discuss this for the field of industrial chemical processes. It is true that there are plenty of data from such processes that are barely used. With the advent of cheap sensors, the amount of such data will rapidly grow in the future. For the present discussion, the question is, if and how such data can be used in the context of thermodynamic modeling or process modeling. The fact that plenty of data are available does not mean that they represent a large range of states. The operators of a chemical plant usually try to keep the plant at the same operating point, so that there are lots of data for a very small range of conditions. Data outside that range, e.g. from start-up or shut-down may be difficult to use for the above-mentioned purposes. Furthermore, the quality of the data may be questionable. E.g., data that are only taken for assessing stationarity or stability need not be accurate.

The discussion basically comes down to the question what the model that is to be developed should accomplish. If it is there to describe the existing operation, data on that operation will prove highly valuable. If it is needed to make predictions, the data on the operation point may be welcome as add-on, but other types of data will be needed for the modeling, and physical models are preferred as compared to mere correlations, cf. also the discussion in Section 4.1.

### **5.2 New sensors**

New sensors are being developed, some of which give new insights in chemical processes. We only discuss medium field NMR spectroscopy as an example here, which is likely to be introduced in the process industry in the coming years [42]. NMR spectroscopy is an analytical tool with a wide range of applications in chemistry, biology, and chemical engineering [43]. It has a high chemical resolution and enables to resolve even chemically similar substances. As it is a non-invasive analysis method, complex fluid mixtures can be investigated without having to take samples. A main advantage over other spectroscopic methods is that NMR spectra can be evaluated quantitatively without prior calibration [44]. Thus, NMR spectroscopy is a particularly attractive method to elucidate and monitor reactions and processes [45]–[49].



In the recent years, medium field NMR spectrometers (also called bench-top NMR spectrometers) have become commercially available. They employ small permanent magnets rather than the superconducting magnets cooled by liquid helium that are used in high-field NMR spectrometers. The medium field NMR spectrometers are therefore compact, robust, comparatively inexpensive, and suited for routine applications in laboratories and production. They also can be used for on-line process monitoring. An example is given in Figure 5, which shows results of a single-stage batch distillation in which the composition of the residue was continuously monitored with medium field NMR spectroscopy [50]. The mixture that was separated was acetonitrile + ethyl formate + dimethyl sulfoxide. The distillation was carried out at 0.1 MPa. The analytical results are plotted in Figure 5 as a function of the temperature, which is also measured. For comparison, also results from a conventional off-line gas chromatographic analysis are shown. After a period of heating-up, the evaporation starts and the concentration of the residue changes. The residue composition measured by on-line NMR spectroscopy is in good agreement with the results obtained from off-line gas chromatography. The new technique lends itself for automated studies of residue curve maps and can be used also in cases where only small amounts of material are available. This opens new perspectives for conceptual process design.

New methods have become available not only for carrying out the NMR experiments but also for the evaluation of the spectra. They include applications of Bayesian statistics [51] and indirect hard modeling [52]. Spectra from series of experiments as the one shown in Figure 5 need to be evaluated automatically. Techniques are presently being developed that make use of the fact that the spectra of the series stem from the same experiment and are hence connected. This shows how the classical experiment and its modeling and simulation merge.

((Figure 5))

## **6. Handling the unknown**

### **6.1. Uncertainties in the thermodynamic model**

The quality of process simulation depends on the quality of the underlying thermodynamic model. The quality of the thermodynamic model, in turn, depends on the quality of experimental data to which the model was fitted. The propagation of the uncertainty of the thermodynamic model into the process model has been discussed by many authors in the literature and many approaches have been described for its assessment, some examples are [53]–[55]. In most approaches the sensitivity is calculated from a variation of the parameters of the thermodynamic model (e.g. the binary interaction parameters of a  $G^E$ -model). From a practical standpoint, it is more attractive to base the sensitivity analysis on a variation of measurable thermo-physical properties of which uncertainties are known as it has been described in [56], [57]. These authors use approaches based on perturbation schemes that

can be implemented in commercial process simulation software in a straightforward manner. The method presented in [57] has been implemented in CHEMASIM, the process simulation software from BASF, as a part of a larger effort by the capabilities of that software were extended regarding sensitivity studies [16] and optimization under uncertainties [17], [18]. An example for the application of this approach is given in Figure 6 in which results for a process are shown, in which CO<sub>2</sub> is removed from flue gas by physical absorption with methanol [57]. The gas solubility is modeled with the SRK equation of state [58], cf. Figure 6, left. Neither the experimental data nor their description with the model are perfect. This is taken into account in the process simulation by perturbing the liquid fugacity coefficient of CO<sub>2</sub>, in the example shown in Figure 6, left, by  $\pm 10\%$ . The impact of the uncertainty of the data and the thermodynamic model on the performance of the absorption process can thereby be assessed, cf. Figure 6, right. This example shows that the analysis of the influence of uncertainties of the thermodynamic model on process simulation results can be routinely considered in the practical process design workflow.

((Figure 6))

## 6.2. Poor specifications

In process design it is usually assumed that the composition of the mixtures in the process is fully specified. Side components are either included in this scheme or neglected and treated only qualitatively. However, this paradigm has limits. It is for example practically impossible to fully specify crude oil and the same holds for many mixtures in biotechnological and polymer processes. Methods have been developed to handle this challenge, the most prominent of which is to use pseudo components, see e.g. [59]–[62]. In some cases also continuous thermodynamics can be applied, see e.g. [60], [63]. Machine learning has also been used in this field [64]–[73].

In the following, a new approach in this field is briefly described which is called NEAT (NMR spectroscopy for Estimation of the Activity of Target components) [74]. It solves the problem of calculating the activity of a known species in a mixture with an arbitrary number of unknown species. NMR spectroscopy is used to quantify the groups that are present in the mixture. Hereby, it is not necessary to identify the unknown species. The group composition that is determined by NMR spectroscopy is mapped to UNIFAC groups and the activity of the target component is calculated. The method has been successfully tested on many mixtures [74]. Some examples are presented in Figure 7, where results are shown for three test systems consisting of a target component (ethanol), water and an additional component that is treated as unspecified within NEAT (i.e. no a priori information on the third component was used). NEAT predicts the influence of the unknown components on the activity coefficient of the target component in the mixtures very well. For more examples, see [74].

((Figure 7))

The approach can be extended: other spectroscopic methods can be used for determining the group composition, and it can be coupled to other thermodynamic group-contribution methods. It is attractive to use NEAT with medium field NMR spectroscopy, cf. Section 5.2. Combined with these instruments, NEAT could also be applied in process analytics. This is another example for how experiments, modeling and simulation merge.

## 7. A preliminary look into machine learning

In machine learning, computers are given access to data from which they “learn” in such a way that they acquire the ability to answer questions regarding situations that resemble those that were studied when taking the data used for the training. Questions that can be answered by machine learning include pattern recognition, classifications, and quantitative statements on outcomes of experiments. There is no clear borderline between training a machine learning algorithm for carrying out quantitative predictions and the fitting of parameters of a flexible correlation. In its pure form, machine learning is entirely mathematical. Hence, adapting the definition given in Section 4.1, it can be considered as a way of developing a correlation, not a model. Still, even establishing such a mathematical correlation of physical facts requires a great deal of physical knowledge: Input and output variables need to be defined, training sets need to be selected, physical data may have to be processed before they can be used in the training set, and the outcome of the training needs to be evaluated.

Neural networks have been applied to thermodynamic problems for more than 20 years [6]. The success was limited, and the initially high interest declined temporarily before surging again after about 2010. Recent applications of machine learning in thermodynamics include solubility or phase equilibrium [75]–[87], thermal (pV<sub>T</sub>) properties [88], [89], caloric properties [90], [91], transport properties [90], [92]–[97], and surface tension [98]–[100], to cite only a few. A substantial part of the recent work is dedicated to properties of ionic liquids [82], [86], [96], [97], [99], [101] that are hard to describe otherwise. Machine learning has also been used for describing the properties of crude oil, asphaltene and natural gas [64], [65], [68]–[70], [72], [83], [93], [102], [103].

In typical successful applications of machine learning, e.g. for pattern recognition, the number of data points that are available for training is extremely large compared to the number of input variables. That ratio is much less favorable for typical thermodynamic problems. Furthermore, thermodynamic models should enable predictions, far beyond the range in which the model was trained. Consider simple  $G^E$ -models like NRTL or UNIQUAC: they can be regarded as mere correlation tools when they are only applied for describing binary phase equilibria of data sets to which they were fitted. But

they enable predictions, e.g. of multi-component phase equilibria, without any further input. These predictions turn out to be astonishingly good in many cases. This is a major achievement and a consequence of the physical background of the models. Machine learning could not accomplish this. It would need training with multi-component data. The realm of machine learning is interpolation, not extrapolation. Is machine learning, hence, doomed in thermodynamics? We think that it is not, but it needs to be combined in a suitable way with the theory of thermodynamics, i.e. with the physical knowledge that we have acquired over generations. Machine learning should be used where physical modeling cannot go. It should complement physical modeling, not replace it. It may, for instance, be attractive to analyze the deviations between a physical model and experimental data (which arise by definition from effects that are not accounted for by the physical model) with machine learning to discern trends.

Furthermore, applications of machine learning to thermodynamic problems can be driven by computational issues. Finding stationary solutions of complex process flowsheets means solving a large system of coupled non-linear equations. Often the feasible domain of the design variables is not known explicitly. This means that it can be difficult to decide whether the reason for non-convergence is due to numerical difficulties (e.g., starting point far away from the unknown solution) or due to the fact that no solution with the specified design variables exists for physical reasons. Machine-learning methods can help to quantify the feasible domain in the design space, thus avoiding time-consuming manual trial-and-error calculations. To this end an adaptive design-of-experiments scheme has been developed to run the simulations, making the method computationally efficient [104].

We discuss only a simple example here, which is taken from [104]: a partial evaporator, cf. Figure 8a. A binary mixture of ethanol and water with 0.1 mol/mol ethanol enters the evaporator with a molar feed flow rate of 150 kmol/h. The temperature and the pressure in the evaporator are chosen as design variables. The feasible operating range of the evaporator is the area between the dew and bubble curves in the  $p,T$ -diagram, in which two phases co-exist. The task is to determine this operating range. There are different ways to solve this task by classical process simulation. E.g., for each value of  $T$  two flash calculations can be carried out, one with a gas-to-feed ratio close to 0 (resulting in  $p^{\text{boil}}$ ) and a second one in which that ratio is close to 1 (resulting in  $p^{\text{dew}}$ ). Any chosen value of  $p$  can then be compared to  $p^{\text{boil}}$  and  $p^{\text{dew}}$ , respectively. This has to be repeated for different values of  $T$ , e.g. from a grid. All this is perfectly feasible for the evaporator that is studied here. In more complex situations corresponding algorithms could be more difficult to devise and their execution could suffer from convergence problems. As an example for machine learning, an adaptive exploration scheme was applied in [104] to learn this feasible range successively from simulations. The procedure commences with a small number of initial design points, as shown in Figure 8b, which are labeled according to the corresponding simulation outcome as either a solution or no-solution. Next, a support vector machine classifier is trained to predict the boundary between the feasible and the infeasible range, cf. Figure

8c. According to this prediction, new design points are placed in regions with high prediction uncertainty, typically at the estimated feasibility range boundary, cf. Figure 8d. The sampling procedure terminates once the progress in learning the boundary satisfies a stopping criterion; the final result for the case considered here is shown in Figure 8e. It can be seen in Figure 8f that the average distance between the learned and the true boundary decreases rapidly, so that for this example not more than 100 simulation points are needed to obtain a reasonable accuracy. The main advantage of the method becomes apparent when comparing it to a uniform sampling scheme where the design points are placed on a uniform grid in the  $p$ - $T$ -plane. At least 400 points would be needed to obtain a comparable accuracy with such a grid.

((Figure 8))

## 8. Conclusions

Digitalization has affected thermodynamics ever since computers became available and computational power has opened many new routes in thermodynamics, e.g. in molecular modeling and simulation. The key topic of digitalization is data and how data are used in modeling and simulation. This is also a key topic in thermodynamics and the currently high momentum in digitalization affects thermodynamics in many ways. This does not concern thermodynamic theory, which stands like a rock, but the ways in which that theory is used in modeling, simulation, and experiments. New connections between these domains are established by digitalization and borders that were once clear begin to vanish. Data from computer simulations can be combined with experimental data into hybrid data sets and used for the development of models of thermodynamic properties of real substances. Conflicting objectives in the development of thermodynamic models can be handled with multi-criteria optimization. New sensors yield new types and increasing amounts of experimental data that can be stored and used. The uncertainty of the experimental data can be considered routinely not only in the development of thermodynamic models but also in their application in process simulation.

Generally, much more data than in the past will be available in the future. But quantity is only one aspect, if it is too high, it may even cause problems. The quality of the data and its usefulness for solving the studied problem are as essential as they always were.

As a consequence of the rapidly growing complexity and interconnectivity of modeling, simulation, and experiments, many tasks in process design can no longer be accomplished by generalists. Specialists will be needed to exploit the new opportunities. They must receive training, which opens new opportunities for industrial – academic co-operations. The complexity also leads to opacity. Even today, it is often practically impossible to fully understand how certain simulation results were obtained. This opacity impedes the assessment of the results and, as a consequence, their application. Such problems will grow in the future. Furthermore, with increasing complexity, the problem of

simulation errors increases. We will have to learn to assess and mitigate them - as we have learned to handle experimental errors.

Machine learning has so far mainly been used in thermodynamics to establish correlations between thermodynamic input and output data sets. As such, it is closely related to parameter fitting, albeit with very flexible functional forms that are parameterized. In most thermodynamic applications the number of data points that are available for training machine-learning algorithms is comparatively small, which limits their success. Machine learning is about interpolating data, not about extrapolating. It is therefore highly attractive to combine machine learning techniques with thermodynamic models with proven predictive capabilities. Machine learning could e.g. help analyzing residues between such models and experimental data for influences that cannot be accounted for by the physical model. We should continue using the broad and deep thermodynamic knowledge that was acquired over centuries and has been used so successfully in physical modeling. But new ways of tackling the unknown with data-driven approaches of machine learning emerge. It will be interesting to see how they can be combined with physical modeling.

### **Acknowledgments**

The authors are grateful for valuable contributions of Anne Friebel, Fabian Jirasek, Andreas Köster, Raoul Heese and Michal Walczak. H.H. gratefully acknowledges support from DFG in a Reinhart Koselleck Project.

### **Symbols used**

$\delta_p$	[%]	deviation in vapor pressure from experimental data
$\delta_\rho$	[%]	deviation in saturated liquid density from experimental data
$\Delta\rho$	[%]	deviation in density from arithmetic mean value
$T$	[K]	absolute temperature
$x_i$	[mol/mol]	mole fraction of component i
$p$	[MPa]	absolute pressure
$\gamma_T$	[-]	activity coefficient of component T
$w$	[m/s]	speed of sound

### **Abbreviations**

GC	gas chromatography
$G^E$	excess Gibbs energy
MCO	multi-criteria optimization
NEAT	NMR spectroscopic method for estimating activity coefficients of known target components in poorly specified mixtures
NMR	nuclear magnetic resonance

NRTL	non-random two liquid
SRK	Soave-Redlich-Kwong
PC(P)-SAFT	perturbed-chain (polar) statistical associating fluid theory
UNIFAC	universal quasi-chemical functional group activity coefficient
UNIQUAC	universal quasi-chemical

## References:

- [1] World Economic Forum, *Digital Transformation of Industries: Societal Implications*, White Paper, **2016**.
- [2] R. Evangelista, P. Guerrieri, V. Meliciani, *Econ. Innov. New Technol.* **2014**, 23 (8), 802. DOI: 10.1080/10438599.2014.918438.
- [3] K. Schwab, *The Fourth Industrial Revolution*, World Economic Forum, **2016**.
- [4] J. M. Prausnitz, T. F. Anderson, E. A. Greens, C. A. Eckert, R. Hsieh, J. P. O’Connell, *Computer calculations of multi-component vapor-liquid and liquid-liquid equilibria.* , Prentice Hall Inc., New Jersey **1980**.
- [5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **1953**, 21 (6), 1087. DOI: 10.1063/1.1699114.
- [6] R. Sharma, D. Singhal, R. Ghosh, and A. Dwivedi, *Comput. Chem. Eng.* **1999**, 23 (3), 385. DOI: 10.1016/S0098-1354(98)00281-6.
- [7] M. Bortz, J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, A. Scheithauer, R. Welke, K.-H. Küfer, H. Hasse, *Comput. Chem. Eng.* **2014**, 60, 354. DOI: 10.1016/j.compchemeng.2013.09.015.
- [8] E. Forte, J. Burger, K. Langenbach, H. Hasse, M. Bortz, *AIChE J.* **2018**, 64 (1), 226. DOI: 10.1002/aic.15857
- [9] E. Forte, University of Kaiserslautern: private communication, **2018**.
- [10] K. Stöbener, P. Klein, S. Reiser, M. Horsch, K.-H. Küfer, H. Hasse, *Fluid Phase Equilib.* **2014**, 373, 100. DOI:10.1016/j.fluid.2014.04.009.
- [11] S. Werth, K. Stöbener, P. Klein, K.-H. Küfer, M. Horsch, H. Hasse, *Chem. Eng. Sci.* **2015**, 121, 110. DOI: 10.1016/j.ces.2014.08.035.
- [12] K. Stöbener, P. Klein, M. Horsch, K.-H. Küfer, and H. Hasse, *Fluid Phase Equilib.* **2016**, 411, 33. DOI:10.1016/j.fluid.2015.11.028.
- [13] J. Burger, N. Asprion, S. Blagov, R. Böttcher, U. Nowak, M. Bortz, R. Welke, K.-H. Küfer, H. Hasse, *Chem. Ing. Tech.* **2014**, 86 (7), 1065. DOI: 10.1002/cite.201400008.
- [14] M. Bortz, V. Maag, J. Schwientek, R. Benfer, R. Böttcher, J. Burger, E. von Harbou, N. Asprion, K.-H. Küfer, H. Hasse, “Computer Aided Chemical Engineering, 37,” in *12<sup>th</sup> International Symposium on Process Systems Engineering and 25<sup>th</sup> European Symposium on Computer Aided Process Engineering* (Eds: K. V. Gernaey, J. K. Huusom, R. Gani) Elsevier, 2063–2068, Copenhagen **2015**.
- [15] J. Burger, V. Papaioannou, S. Gopinath, G. Jackson, A. Galindo, C. S. Adjiman, *AIChE J.* **2015**, 61 (10), 3249. DOI: 10.1002/aic.14838.
- [16] N. Asprion *et al.*, “INES – An Interface Between Experiments and Simulation to Support the Development of Robust Process Designs,” *Chem Ing Tech*, vol. 87, no. 12, pp. 1810–1825, 2015.
- [17] M. Bortz *et al.*, “Efficient Approach for Calculating Pareto Boundaries under Uncertainties in Chemical Process Design,” *Ind. Eng. Chem. Res.*, vol. 56, no. 44, pp. 12672–12681, 2017.
- [18] N. Asprion, S. Blagov, R. Böttcher, J. Schwientek, J. Burger, E. von Harbou, M. Bortz, *Chem. Ing. Tech.* **2017**, 89 (5), 665. DOI: 10.1002/cite.201600098.
- [19] E. Forte, E. von Harbou, J. Burger, N. Asprion, M. Bortz, *Chem Ing Tech* **2017**, 89 (5), 645. DOI: 10.1002/cite.201600104.
- [20] J. Gross, G. Sadowski, *Fluid Phase Equilib* **2000**, 168 (2), 183. DOI: 10.1016/S0378-3812(00)00302-2.

- [21] J. Gross, G. Sadowski, *Ind Eng Chem Res* **2001**, *40* (4), 1244. DOI: 10.1021/ie0003887.
- [22] J. Gross, J. Vrabec, *AIChE J.* **2006**, *52* (3), 1194. DOI: 10.1002/aic.10683.
- [23] D. R. Lide, *CRC Handbook of Chemistry and Physics*, 86th ed. Boca Raton: CRC Press/Taylor and Francis, **2005**.
- [24] M. Thol, G. Rutkai, A. Köster, M. Kortmann, R. Span, J. Vrabec, *Chem. Eng. Sci.* **2015**, *121*, 87. DOI: 10.1016/j.ces.2014.07.051.
- [25] M. Thol, F. H. Dubberke, G. Rutkai, T. Windmann, A. Köster, R. Span, J. Vrabec, *Fluid Phase Equilibria* **2016**, *418*, 133. DOI: 10.1016/j.fluid.2015.09.047.
- [26] M. Thol, G. Rutkai, A. Köster, F. H. Dubberke, T. Windmann, R. Span, J. Vrabec, *J. Chem. Eng. Data* **2016**, *61* (7), 2580. DOI: 10.1021/acs.jced.6b00261.
- [27] M. Thol, G. Rutkai, A. Köster, S. Miroshnichenko, W. Wagner, J. Vrabec, R. Span, *Mol. Phys.* **2017**, *115* (9-12), 1166. DOI: 10.1080/00268976.2016.1262557.
- [28] R. Lustig, *Mol. Simul.* **2011**, *37* (6), 457. DOI: 10.1080/08927022.2011.552244.
- [29] R. Lustig, *Mol. Phys.* **2012**, *110* (24), 3041. DOI: 10.1080/00268976.2012.695032.
- [30] C. W. Glass, S. Reiser, G. Rutkai, S. Deublein, A. Köster, G. Guevara-Carrion, A. Wafai, M. Horsch, M. Bernreuther, T. Windmann, H. Hasse, J. Vrabec, *Comput. Phys. Commun.* **2014**, *185* (12), 3302. DOI: 10.1016/j.cpc.2014.07.012.
- [31] G. Rutkai, A. Köster, G. Guevara-Carrion, T. Janzen, M. Schappals, C. W. Glass, M. Bernreuther, A. Wafai, S. Stephan, M. Kohns, S. Reiser, S. Deublein, M. Horsch, H. Hasse, J. Vrabec, *Comput. Phys. Commun.* **2017**, *221*, 343. DOI: 10.1016/j.cpc.2017.07.025.
- [32] F. H. Case, J. Brennan, A. Chaka, K. D. Dobbs, D. G. Friend, P. A. Gordon, J. D. Moore, R. D. Mountain, J. D. Olson, R. B. Ross, M. Schiller, V. K. Shen, E. A. Stahlberg, *Fluid Phase Equilibria* **2008**, *274* (1), 2. DOI: 10.1016/j.fluid.2008.05.009.
- [33] B. Eckl, J. Vrabec, H. Hasse, *Fluid Phase Equilibria* **2008**, *274* (1), 16. DOI: 10.1016/j.fluid.2008.02.002.
- [34] P. Colonna, N. R. Nannan, A. Guardone, E. W. Lemmon, *Fluid Phase Equilibria* **2006**, *244* (2), 193. DOI: 10.1016/j.fluid.2006.04.015.
- [35] G. Rutkai, J. Vrabec, *J. Chem. Eng. Data* **2015**, *60* (10), 2895. DOI: 10.1021/acs.jced.5b00266.
- [36] A. Köster, T. Jiang, G. Rutkai, C. W. Glass, J. Vrabec, *Fluid Phase Equilibria* **2016**, *425*, 84. DOI: 10.1016/j.fluid.2016.05.002.
- [37] M. P. Allen, D. J. Tildesley, *Computer simulation of liquids*. Clarendon Press, Oxford **1989**.
- [38] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, vol. 2nd ed., Academic Press, San Diego **2002**.
- [39] B. N. Taylor, C. E. Kuyatt, "Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results," National Institute of Standards and Technology, Physics Laboratory, Gaithersburg, NIST Technical Note 1297, **1993**.
- [40] W. F. van Gunsteren, A. E. Mark, *J. Chem. Phys.* **1998**, *108* (15), 6109. DOI: 10.1063/1.476021.
- [41] M. Schappals, A. Mecklenfeld, L. Kröger, V. Botan, A. Köster, S. Stephan, E. J. García, G. Rutkai, G. Raabe, P. Klein, K. Leonhard, C. W. Glass, J. Lenhard, J. Vrabec, H. Hasse, *J. Chem. Theory Comput.* **2017**, *13* (9), 4270. DOI: 10.1021/acs.jctc.7b00489.
- [42] K. Meyer, S. Kern, N. Zientek, G. Guthausen, M. Maiwald, *TrAC Trends Anal. Chem.*, **2016**, *83*, 39. DOI: 10.1016/j.trac.2016.03.016.
- [43] O. Zerbe, S. Jurt, *Applied NMR spectroscopy for chemists and life scientists*. John Wiley & Sons, 2013.
- [44] S. K. Bharti, R. Roy, *TrAC Trends Anal. Chem.*, **2012**, *35*, 5. DOI: 10.1016/j.trac.2012.02.007.
- [45] M. Maiwald, H. H. Fischer, M. Ott, R. Peschla, C. Kuhnert, C. G. Kreiter, G. Maurer, H. Hasse, *Ind. Eng. Chem. Res.* **2003**, *42* (2), 259. DOI: 10.1021/ie0203072.
- [46] M. A. Bernstein, M. Štefinović, C. J. Sleight, *Magn. Reson. Chem.* **2007**, *45* (7), 564. DOI: 10.1002/mrc.2007.
- [47] A. Brächer, S. Hoch, K. Albert, H. J. Kost, B. Werner, E. von Harbou, H. Hasse, *J. Magn. Reson.* **2014**, *242*, 155. DOI: 10.1016/j.jmr.2014.02.013.
- [48] É. J. Kibrik, O. Steinhof, G. Scherr, W. R. Thiel, H. Hasse, *Ind. Eng. Chem. Res.* **2014**, *53* (32), 12602. DOI: 10.1021/ie5001746.
- [49] A. Scheithauer, E. von Harbou, H. Hasse, T. Grützner, C. Rijksen, D. Zollinger, W. R. Thiel, *AIChE J.* **2015**, *61* (1), 177. DOI: 10.1002/aic.14623.



- [50] A. Friebel, University of Kaiserslautern: private communication, **2018**.
- [51] Y. Matviychuk, E. von Harbou, D. J. Holland, *J. Magn. Reson.* **2017**, 285, 86. DOI: 10.1016/j.jmr.2017.10.009.
- [52] A. Michalik-Onichimowska, S. Kern, J. Riedel, U. Panne, R. King, M. Maiwald, *J. Magn. Reson.* **2017**, 277, 154. DOI: 10.1016/j.jmr.2017.02.018.
- [53] S. Hajipour, M. A. Satyro, M. W. Foley, *Energy Fuels* **2014**, 28, (2), 1569. DOI: 10.1021/ef4019838
- [54] W. B. Whiting, *J. Chem. Eng. Data* **1996**, 41 (5), 935. DOI: 10.1021/je9600764.
- [55] Y. Xin, W. B. Whiting, *Ind. Eng. Chem. Res.* **2000**, 39 (8), 2998. DOI: 10.1021/ie990617t.
- [56] P. M. Mathias, *J. Chem. Eng. Data* **2014**, 59 (4), 1006. DOI: 10.1021/je400748p.
- [57] J. Burger, N. Aspiron, S. Blagov, M. Bortz, *J. Chem. Eng. Data* **2017**, 62 (1), 268. DOI: 10.1021/acs.jced.6b00633.
- [58] G. Soave, *Chem Eng Sci* **1972**, 27 (6), 1197. DOI: 10.1016/0009-2509(72)80096-4.
- [59] A. Danesh, *PVT and phase behaviour of petroleum reservoir fluids*, vol. 47. Elsevier, **1998**.
- [60] D. Browarzik, "Chapter 9: Polydisperse Fluids," in *Applied Thermodynamics of Fluids*, A. R. H. Goodwin, J. V. Sengers, C. J. Peters, Eds. London: Royal Society of Chemistry, **2010**.
- [61] S. Enders and B. A. Wolf, *Polymer thermodynamics: liquid polymer-containing mixtures*, vol. 238. Springer Science & Business Media, **2011**.
- [62] E. Forte, S. E. Taylor, *Adv. Colloid Interface Sci.* **2015**, 217 (1). DOI: 10.1016/j.cis.2014.12.002.
- [63] H. Kehlen, M. T. Rätzsch, J. Bergmann, *AIChE J.* **1985**, 31 (7), 1136. DOI: 10.1002/aic.690310712.
- [64] S. O. Olatunji, A. Selamat, A. A. A. Raheem, *Expert Syst. Appl.* **2011**, 38 (9), 10911. DOI: 10.1016/j.eswa.2011.02.132.
- [65] A. Chamkalani, A. H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, *Chem. Eng. Sci.* **2012**, 81, 202. DOI: 10.1016/j.ces.2012.06.060.
- [66] M. Li, X. Huang, H. Liu, B. Liu, Y. Wu, A. Xiong, T. Dong, *Fluid Phase Equilib.* **2013**, 356, 11. DOI: 10.1016/j.fluid.2013.07.017.
- [67] A. Kamari, A. Khaksar-Manshad, F. Gharagheizi, A. H. Mohammadi, S. Ashoori, *Ind. Eng. Chem. Res.* **2013**, 52 (44), 15664. DOI: 10.1021/ie402462q.
- [68] A. Farasat, A. Shokrollahi, M. Arabloo, F. Gharagheizi, A. H. Mohammadi, *Fuel Process. Technol.* **2013**, 115, 201. DOI: 10.1016/j.fuproc.2013.06.007.
- [69] A. Hemmati-Sarapardeh, R. Alipour-Yeganeh-Marand, A. Naseri, A. Safiabadi, F. Gharagheizi, P. Ilani-Kashkouli, A. H. Mohammadi, *Fluid Phase Equilib.* **2013**, 354, 177. DOI: 10.1016/j.fluid.2013.06.005
- [70] S. Rafiee-Taghanaki, M. Arabloo, A. Chamkalani, M. Amani, M. H. Zargari, M. R. Adelzadeh, *Fluid Phase Equilib.* **2013**, 346, 25. DOI: 10.1016/j.fluid.2013.02.012.
- [71] A. Shokrollahi, M. Arabloo, F. Gharagheizi, A. H. Mohammadi, *Fuel* **2013**, 112, 375. DOI: 10.1016/j.fuel.2013.04.036.
- [72] A. Gholami, M. Asoodeh, P. Bagheripour, *Fluid Phase Equilib.* **2014**, 382, 139. DOI: 10.1016/j.fuel.2013.04.036.
- [73] H. Ziaee, S. M. Hosseini, A. Sharafpoor, M. Fazavi, M. M. Ghiasi, A. Bahadori, *J. Taiwan Inst. Chem. Eng.* **2015**, 46, 205. DOI: 10.1016/j.jtice.2014.09.015.
- [74] F. Jirasek, J. Burger, and H. Hasse, *Ind Eng Chem Res* **2018**, in press. DOI: 10.1021/acs.iecr.8b00917.
- [75] A. Eslamimanesh, F. Gharagheizi, A. H. Mohammadi, D. Richon, *Chem. Eng. Sci.*, **2011**, 66 (13), 3039. DOI: 10.1016/j.ces.2011.03.016.
- [76] A. Eslamimanesh, F. Gharagheizi, M. Illbeigi, A. H. Mohammadi, A. Fazlali, D. Richon, *Fluid Phase Equilib.* **2012**, 316, 34. DOI: 10.1016/j.fluid.2011.11.029.
- [77] A. H. Mohammadi, A. Eslamimanesh, D. Richon, F. Gharagheizi, M. Yazdizadeh, J. Javanmardi, H. Hashemi, M. Zarifi, S. Babaei, *Ind. Eng. Chem. Res.* **2012**, 51 (2), 1062, 2012. DOI: 10.1021/ie201904r.
- [78] M. Mesbah, E. Soroush, A. Shokrollahi, A. Bahadori, *J. Supercrit. Fluids* **2014**, 90, 110. DOI: 10.1016/j.supflu.2014.03.009.
- [79] H. Safari, A. Shokrollahi, A. Moslemizadeh, M. Jamialahmadi, M. H. Ghazanfari, *Fluid Phase Equilib.* **2014**, 374, 86. DOI: 10.1016/j.fluid.2014.04.023.

- [80] A. R. Bahmani, F. Sabzi, M. Bahmani, *J. Mol. Liq.* **2015**, *211*, 395. DOI: 10.1016/j.molliq.2015.07.009.
- [81] M. Mesbah, E. Soroush, V. Azari, M. Lee, A. Bahadori, S. Habibnia, *J. Supercrit. Fluids* **2015**, *97*, 256. DOI: 10.1016/j.supflu.2014.12.011.
- [82] H. R. Amedi, A. Baghban, M. A. Ahmadi, *J. Mol. Liq.* **2016**, *216*, 411. DOI: 10.1016/j.molliq.2016.01.060.
- [83] A. Baghban, S. Namvarrechi, L. T. K. Phung, M. Lee, A. Bahadori, T. Kashiwao, *Pet. Sci. Technol.* **2016**, *34 (16)*, 1431. DOI: 10.1080/10916466.2016.1202966.
- [84] M. Ali Ahmadi, A. Ahmadi, *Int. J. Low-Carbon Technol.*, **2016**, *11 (3)*, 325. DOI: 10.1093/ijlct/ctu034.
- [85] H. Taghvaei, M. A. Amooie, A. Hemmati-Sarapardeh, H. Taghvaei, *J. Mol. Liq.* **2016**, *224*, 745. DOI: 10.1016/j.molliq.2016.09.119.
- [86] A. Baghban, A. H. Mohammadi, M. S. Taleghani, *Int. J. Greenh. Gas Control* **2017**, *58*, 19. DOI: 10.1016/j.ijggc.2016.12.009.
- [87] H. Yarveicy, M. M. Ghiasi, A. H. Mohammadi, *J. Mol. Liq.* **2018**, *255*, 375. DOI: 10.1016/j.molliq.2017.11.156.
- [88] A. Sözen, E. Arcaklioğlu, T. Menlik, M. Özalp, *Expert Syst. Appl.* **2009**, *36 (3)*, 4346. DOI: 10.1016/j.eswa.2008.05.023.
- [89] G. C. Sosso, V. L. Deringer, S. R. Elliott, G. Csányi, *Mol. Simul.* **2018**, *0 (0)*, 1. DOI: 10.1080/08927022.2018.1447107.
- [90] A. Şencan, İ. İ. Köse, R. Selbaş, *Energy Convers. Manag.*, **2011**, *52 (2)*, 958. DOI: 10.1016/j.enconman.2010.08.024.
- [91] A. Kamari, M. Sattari, A. H. Mohammadi, D. Ramjugernath, *Fluid Phase Equilib.* **2016**, *412*, 228. DOI: 10.1016/j.fluid.2015.12.033.
- [92] D. A. Saldana, L. Starck, P. Mougin, B. Rousseau, N. Ferrando, B. Creton, *Energy Fuels* **2012**, *26 (4)*, 2416. DOI: 10.1021/ef3001339.
- [93] M. A. Al-Marhoun, S. Nizamuddin, A. A. Raheem, S. S. Ali, A. A. Muhammadain, *J. Pet. Sci. Eng.* **2012**, *86*, 111. DOI: 10.1016/j.petrol.2012.03.029.
- [94] F. Ghaderi, A. H. Ghaderi, B. Najafi, N. Ghaderi, *J. Supercrit. Fluids* **2013**, *81*, 67. DOI: 10.1016/j.supflu.2013.04.017.
- [95] Y. Zhao, X. Zhang, L. Deng, S. Zhang, *Comput. Chem. Eng.* **2016**, *92*, 37. DOI: 10.1016/j.compchemeng.2016.04.035.
- [96] A. Baghban, M. N. Kardani, S. Habibzadeh, *J. Mol. Liq.* **2017**, *236*, 452. DOI: 10.1016/j.molliq.2017.04.019.
- [97] M.-R. Fatehi, S. Raeissi, D. Mowla, *J. Mol. Liq.* **2017**, *227*, 309. DOI: 10.1016/j.molliq.2016.11.133.
- [98] A. Roosta, P. Setoodeh, A. Jahanmiri, *Ind. Eng. Chem. Res.* **2012**, *51 (1)*, 561. DOI: 10.1021/ie2017459.
- [99] M. Hashemkhani, R. Soleimani, H. Fazeli, M. Lee, A. Bahadori, M. Tavalaeian, *J. Mol. Liq.*, **2015**, *211*, 534. DOI: 10.1016/j.molliq.2015.07.038.
- [100] M. Nabipour, P. Keshavarz, *Int. J. Refrig.* **2017**, *75*, 217. DOI: 10.1016/j.ijrefrig.2016.12.011.
- [101] M.-A. Ahmadi, B. Pouladi, Y. Javvi, S. Alfkhani, R. Soleimani, *J. Supercrit. Fluids* **2015**, *97*, 81. DOI: 10.1016/j.supflu.2014.11.009.
- [102] M. Ghavipour, M. Ghavipour, M. Chitsazan, S. H. Najibi, S. S. Ghidary, *Chem. Eng. Res. Des.* **2013**, *91 (2)*, 264. DOI: 10.1016/j.cherd.2012.08.010.
- [103] M. M. Ghiasi, H. Yarveicy, M. Arabloo, A. H. Mohammadi, R. M. Behbahani, *J. Mol. Liq.* **2016**, *223*, 1081. DOI: 10.1016/j.molliq.2016.09.009.
- [104] M. Bortz, Fraunhofer Institute: private communication, **2018**.
- [105] H. Hasse and J. Lenhard, "Boon and Bane: On the Role of Adjustable Parameters in Simulation Models," in *Mathematics as a Tool: Tracing New Roles of Mathematics in the Sciences*, J. Lenhard and M. Carrier, Eds. Cham: Springer International Publishing, **2017**, pp. 93–115.

**Figure captions:**

**Figure 1:** Classical picture of relations between reality and models, simulation and experiments. The feedback loop of model development is also shown. All blocks are affected by the current wave of digitalization and the borders between the blocks become blurred. Adopted from [105].

**Figure 2:** Example for the application of multi-criteria optimization (MCO) for developing thermodynamic models [8]: Pareto fronts for three different versions of a PC(P)-SAFT equation for water. The two conflicting objectives are deviations between the model results for the vapor pressure  $\delta_p$  and the saturated liquid density  $\delta_\rho$ . The Pareto fronts are represented by symbols (Pareto-optimal solutions) connected by linear splines. The different symbols represent the different PC(P)-SAFT versions. Squares: non-polar; circles: polar with fixed dipole moment from the literature; triangles: polar with dipole moment adjusted.

**Figure 3:** Example for the application of hybrid data sets for the development of equations of state [25]: speed of sound of hexamethyldisiloxane (HDMS) along four isotherms. Open circles: molecular simulation data; filled symbols: experimental data; solid line: equation of state, based hybrid data set from experiments and computer simulations; dotted line: equation of state of Colonna et al. [34]; dashed line: vapor pressure curve.

**Figure 4:** Example from results of a round-robin study [41] that confirms the existence of systematic errors in complex simulations (adapted from [41]). The studied property is the density of *i*-butane at 41 MPa described by the using the OPLS force field. The symbols correspond to data from different groups obtained with different codes (for details, see [41]). Not the primary data are shown but their deviation from the arithmetic mean of all results for a given temperature. They exceed the statistical uncertainties of the individual simulation results which are often below 0.05%.

**Figure 5:** Example for the application of medium field NMR spectroscopy as new sensor for process monitoring [50]: Left: results for the composition of the residue in a single-stage batch distillation of a mixture of acetonitrile, ethyl formate, and dimethyl sulfoxide (DMSO) at  $p = 0.1$  MPa plotted as a function of the temperature. The symbols correspond to experimental data. Open symbols: on-line NMR; filled symbols: off-line GC. Triangles: acetonitrile; squares: ethyl formate; diamonds: dimethyl sulfoxide. Right: typical medium field  $^1\text{H}$  NMR spectrum with peak assignment for the studied mixture.

**Figure 6:** Example for handling uncertainties in experimental data and thermodynamic models using a perturbation scheme: Application to the absorption of  $\text{CO}_2$  from flue gas using methanol as solvent. Left: Overall pressure versus solubility of  $\text{CO}_2$  in methanol for 298 K (circles) and 318 K (squares). The solid lines correspond to the nominal model, the dashed lines to the perturbed model. Right: trade-off between the number of stages and the solvent flow rate needed to remove 99% of the  $\text{CO}_2$ .

The solid line corresponds to the nominal model; the dashed line corresponds to the results from the perturbed model. For details, see [58].

**Figure 7:** Example for handling poor specifications using the NEAT method [74]: Activity coefficient of a target component (T = ethanol) versus mole fraction of an unspecified component (U = acetic acid, methyl acetate or 2-butanone) in ternary mixtures with water at  $T = 298$  K and  $p = 0.1$  MPa for a fixed molar ratio of ethanol to water. The lines correspond to the results from UNIFAC for the completely specified mixture (dashed: acetic acid; dash-dotted: methyl acetate; dotted: 2-butanone). The symbols are predictions with the NEAT method for the unspecified mixture, i.e., using no information on component U (triangles: acetic acid; circles: methyl acetate; squares: 2-butanone).

**Figure 8:** Example for the application of machine learning to solving thermodynamic tasks in process simulation [104]: determination of the feasible operating range of a partial evaporator. a: Sketch of the evaporator; b: initial set of design points to start the exploration of the feasible range; c-e: successive exploration of the feasible domain; f: average distance between the true and the learned boundaries separating the feasible domain (i.e. two-phase coexistence) from the infeasible (single phase) domains. In plots b-e the dashed lines are the true dew and bubble curves; red points encode design choices resulting in a solution whereas for blue points no solution could be found.

### **Table of contents:**

The influence of the current wave of digitalization on thermodynamics is analyzed. Large amounts of data become easily available. The power and complexity of thermodynamic models and simulation techniques is rapidly increasing, and new routes become viable to link them to the data. Machine learning opens new perspectives.

Fig.1

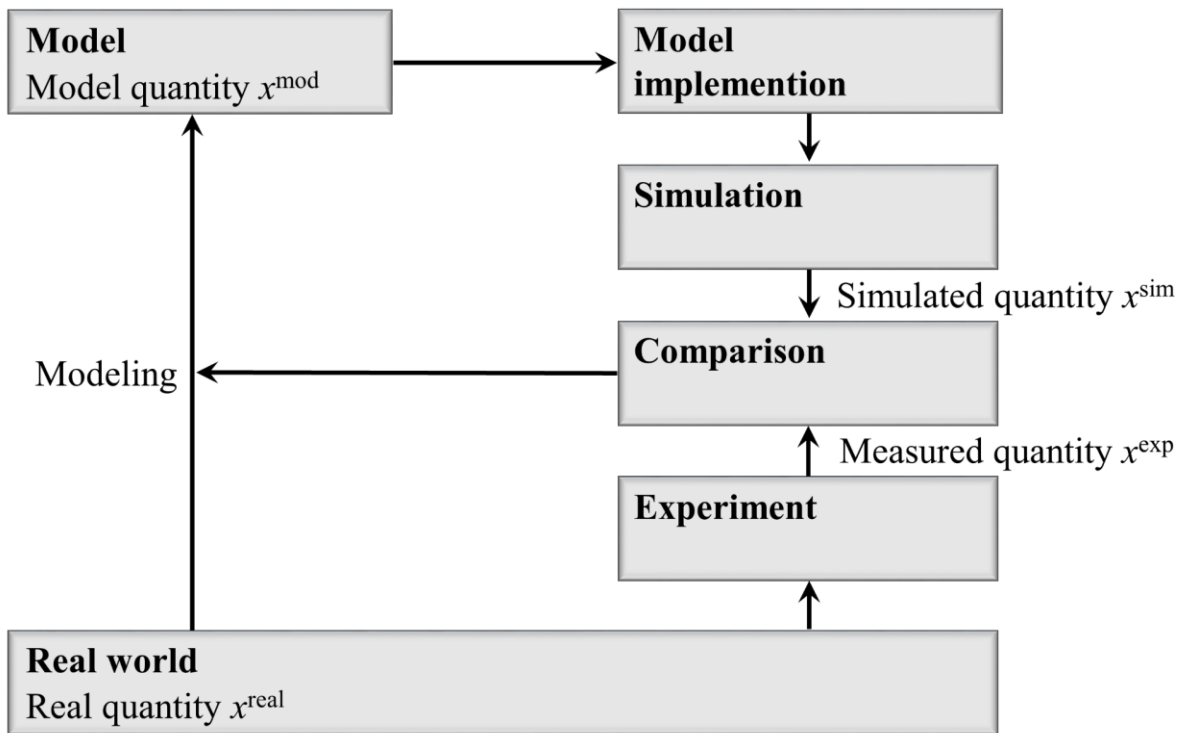


Fig. 2

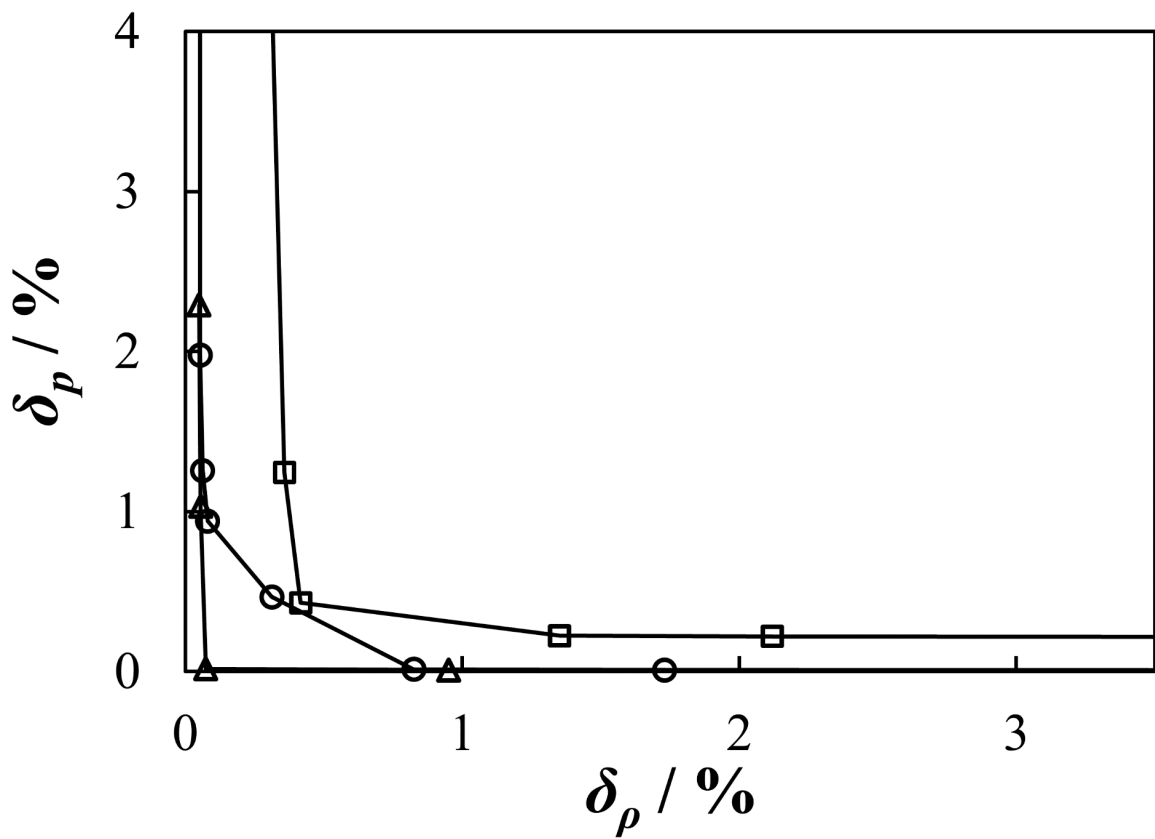


Fig. 3

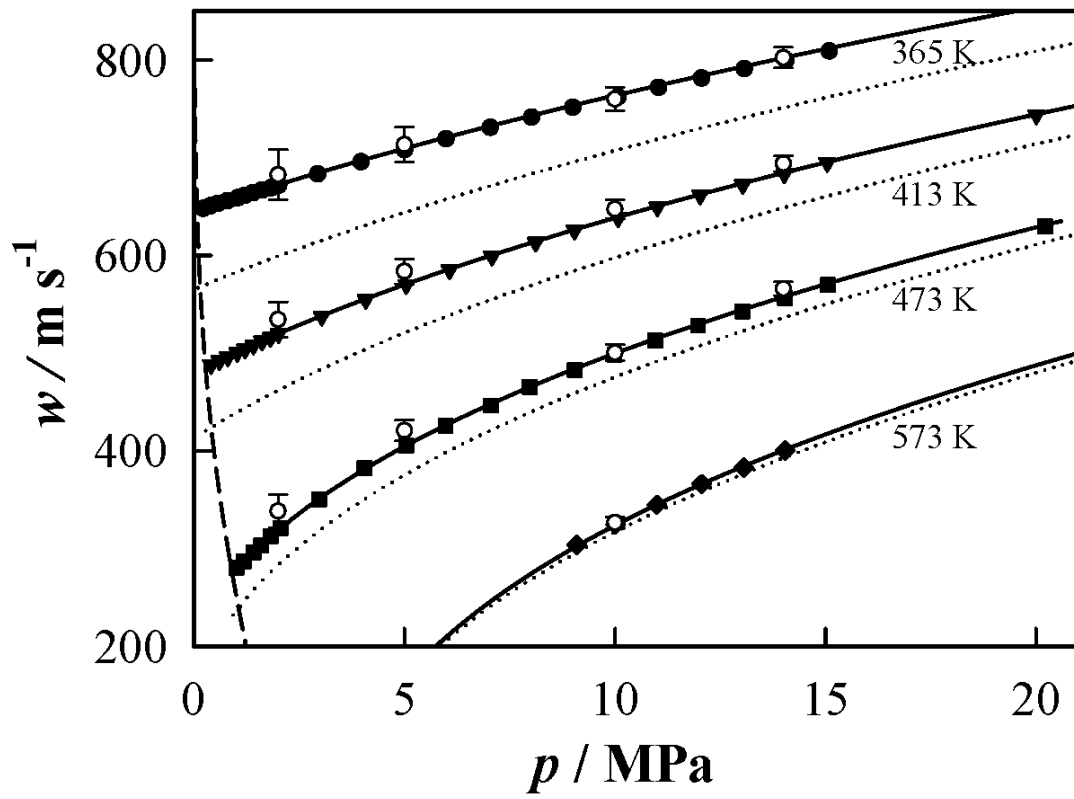


Fig. 4

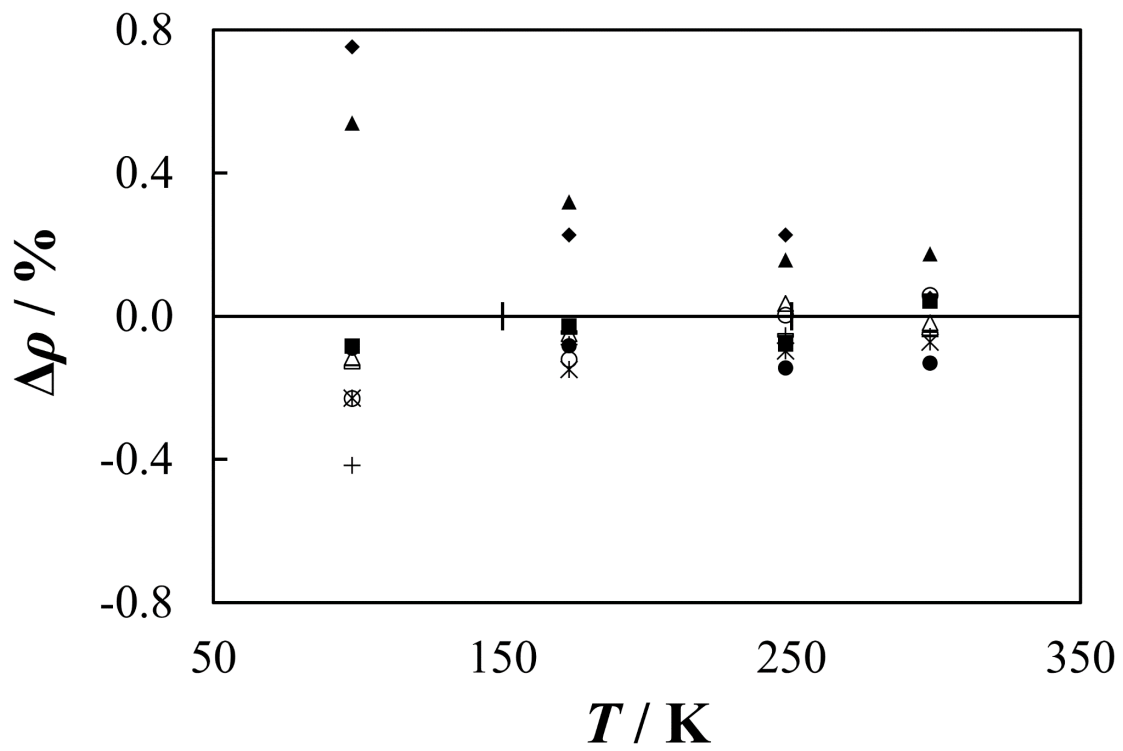


Fig. 5

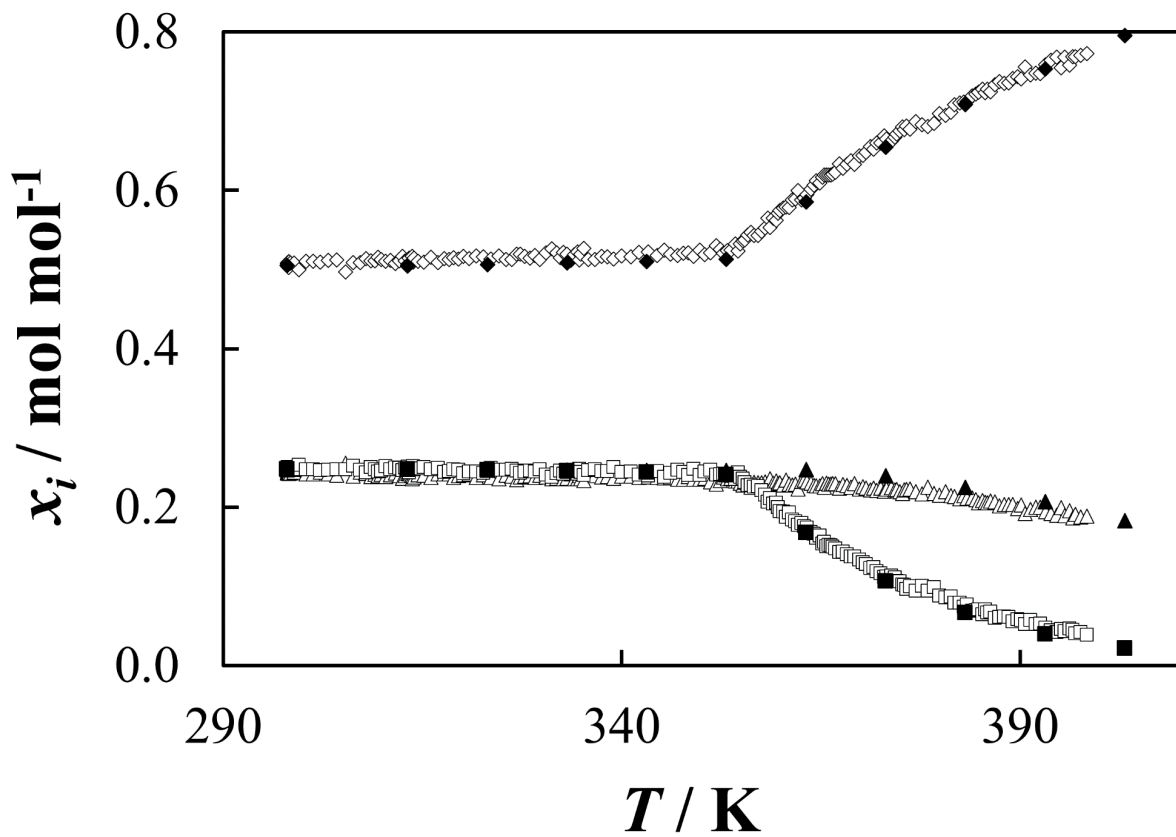


Fig. 6

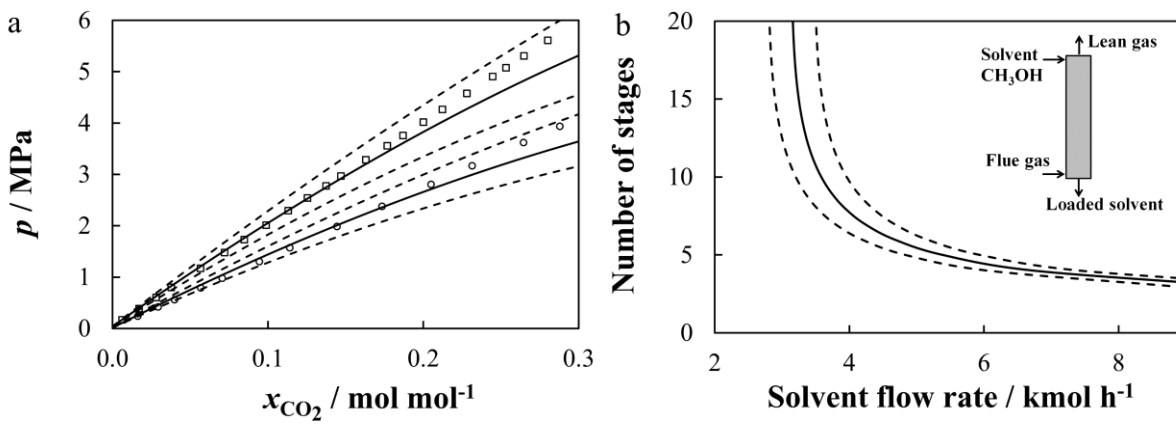


Fig. 7

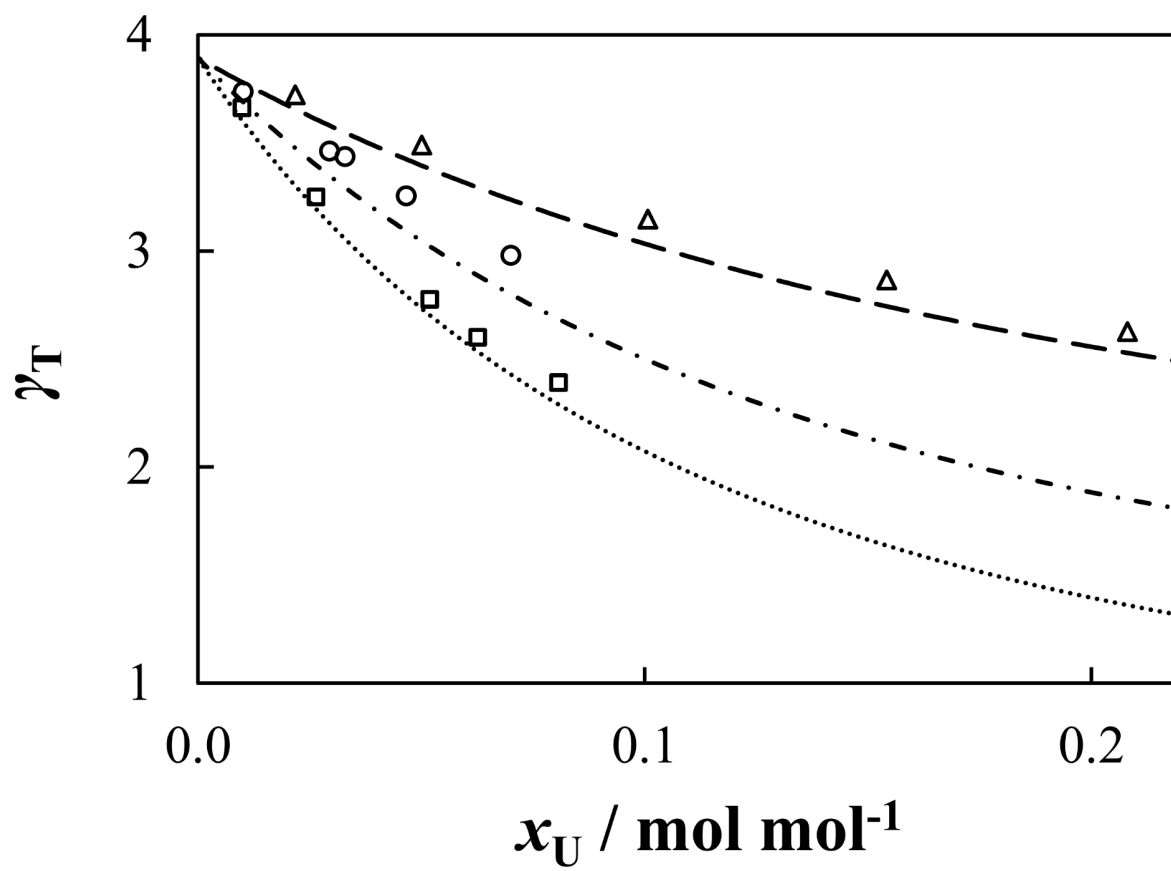




Fig. 8

