Data quality and assessment, validation methods and error propagation through the simulation software: Report from the Round-Table Discussion at the 10th World Congress of Chemical Engineering in Barcelona (October 1–5, 2017)

P. Mathias, A. Soto, L. Fele-Zilnik, J.-C de Hemptinne, A. Bazyleva, J. Abildskov

Data Quality and Assessment, Validation Methods and Error Propagation through the Simulation Software:

Report from the Round-Table Discussion at the 10th World Congress of Chemical Engineering in Barcelona (October 1-5, 2017)

P.M. Mathias[(1)]; A. Soto[(2)]; L. Fele-Zilnik[(3)]; J.-C. de Hemptinne[(4)]; A. Bazyleva[(5)]; J. Abildskov[(6)]

1 : Fluor Corporation, 3 Polaris Way, Aliso Viejo, California 92698, United States
2 : Department of Chemical Engineering, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain
3 : Department of Catalysis and Chemical Reaction Engineering, National Institute of Chemistry, Hajdrihova 19, P.O.Box 660, SI-1001 Ljubljana, Slovenia
4 : IFP Energies nouvelles, 1&4 Avenue de Bois Préau, 92852 Rueil-Malmaison Cedex, France
5 : Applied Chemicals and Materials Division, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305-3337, USA
6 : PROSYS, Department of Chemical and Biochemical Engineering, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Abstract:

The issues of data quality and propagation of data uncertainties into process design and plant specifications are of great current interest. Hence, two Working Parties of the European Federation of Chemical Engineers (EFCE) organized a Round Table Discussion on the topic, as part of the World Congress of Chemical Engineering (WCCE10) in Barcelona, in October 2017. The discussion was guided by industrial and academic experts, with the audience as a key part of the discussion, trying to find some answers in three areas: Data acquisition and evaluation of experimental uncertainties, tools for data reconciliation to improve their quality, and impact of data uncertainties on the process at the end.

Several concrete stories are presented that demonstrate the importance of considering data quality and all possible contributions to the uncertainty of chemical process design. Difficulties associated with data quality are discussed at various levels: (1) the experimentalists (measurement issues, evaluation of uncertainties, use of consistency analysis tools); (2) model developers (capture of adequate physics, parameter regression strategies, uncertainty propagation), (3) vendors of process simulation software, and (4) process engineers (who are responsible at the end).

Paths for improvements were proposed through better and more efficient communication among different participants, as well as through education.

Keywords:

data quality and assessment, uncertainties, data reconciliation, error propagation, process simulation

Introduction (context)

Uncertainty assessment and analysis as well as the impact of the uncertainty concept on thermodynamic property modeling have become central issues for scientists and researchers in the field of thermodynamics and chemical engineering lately [1,2]. These uncertainties result from a combination of factors, starting from data, through the use of a thermodynamic model, all the way to models for process units in a process simulator.

Recently, there has been an increased focus on the quality of measured data, on the recognition and quantification of errors, where they occur, how they are further transmitted and what their impact might be, not only on thermodynamic property calculations, but also on process simulations [3]. The uncertainty concept has not received the deserved attention of process engineers in the past. Several research studies evaluating the impact of uncertainties in property models on process design and plant operability were performed in the past with different outcomes [4,5] [6,7] and references 43-48 in [7]. In some cases, rough approximations of physical properties are good enough for process design, but sometimes extreme sensitivity of design to physical properties is observed. As stressed by Mathias [8], uncertainty analysis is still not a routine task of today's industrial practice. Mathias [8] suggested that this is mainly due to the lack of education and awareness of process engineers and due to the fact that the methods are quite difficult to apply. It is concluded in [1] that two barriers must be overcome in order to implement uncertainty analysis in everyday chemical engineering practice: education on uncertainties should be introduced in undergraduate and graduate courses, and easy to use methods should become available in process simulators.

These issues were highlighted during a Round Table Discussion on the topic 'Data Quality and Assessment, Validation Methods and Error Propagation through the Simulation Software' that was organized during the 10th World Congress of Chemical Engineering (WCCE10) in Barcelona (October 1st-5th 2017) as a joint action between two Working Parties of the European Federation of Chemical Engineers (EFCE), namely the Working Party on Thermodynamics and Transport Properties and the Working Party on Fluid Separations. The organizers of the Round Table were A. Soto (Spain), L. Fele-Zilnik (Slovenia), J.-C. de Hemptinne (France). Paul M. Mathias (USA) served as the chair of the discussion panel. Six panelists from both Working Parties were asked to contribute to the discussion with some suggested topics given at the beginning of the Round Table. The content of the discussion, the conclusions, and main messages are presented here.

Three questions were in the center of the Round Table discussion:

- Data acquisition: How are experimental uncertainties evaluated?
- Data reconciliation: What tools can be used to improve the quality of existing data (e.g. consistency analysis, use of predictive models)?
- Impact of data on process simulation: What data have the most impact on the process, and how can their effects be quantified?

The purpose of this document is to report on the main lessons learned from the discussion. It is presented in three steps: first with some examples showing the importance of the issue, second by describing the difficulties encountered by the various types of actors (experimentalist, process engineers, *etc*.), and finally by suggesting some paths forward. The conclusion summarizes the outcomes.

Some Concrete Stories

Several examples which demonstrate the importance of considering data quality and all possible contributions to the uncertainty of process design were presented.

An example of a wrong decision made because of a combination of insufficient or wrong phase equilibrium data with improper interpretation of column parameters, was the explosion at a butadiene distillation facility in Texas City in 1969 [9–11]. In order to repair the stripper compressor, the distillation unit was placed on total reflux. However, the column was slowly losing material because of a leaking valve in the overhead line. Vinylacetylene is the most dangerous impurity to be separated in the process, since it becomes explosive above a certain concentration. Based on the fact that it has the highest boiling temperature among all components in the distillation column, its concentration was monitored at the column base and was shown to be below the hazardous level, while aftermath modeling showed non-ideal mixture behavior with the highest concentration expected higher up, between 10^{th} and 15^{th} trays. That is exactly where the explosion happened.

One panelist provided another example, although references were not provided [12]. This incident related to the release of gas from a high pressure (80 MPa) gas condensate reservoir. Such a release follows the Joule-Thomson effect. The process simulator calculated a temperature drop, while in practice a temperature increase was observed. The main reason for such dramatic disagreement was that parameters of an equation of state (EOS) were fitted to a wrong data set.

Also, when comparing working fluids for thermodynamic cycles, the net power output is the key relevant quantity. When one takes into consideration its 0.95 confidence interval, for each fluid, additional quantitative information becomes vital for the fluid selection. The ranking of working fluids can be significantly different based on whether the mean value of the net power output is used as a criterion, or, alternatively, whether uncertainties (*e.g.*, the lower bound of the 0.95 confidence interval) are incorporated [13].

A final example mentioned by the panel concerns the use of temperature-independent binary interaction parameters for vapour-liquid equilibrium (VLE) computations in the design of a pressure swing absorption tower. This may lead to significant errors in the estimation of the number of theoretical stages of the column.

The panel concluded that all these examples clearly demonstrate the effect of uncertainty on the quality of the design and the ability to anticipate unsafe plant operation.

Difficulties at different levels:

- **For experimentalists: Property measurements**

In spite of very rich existing databases, there is still a lack of property data for a large number of chemical systems of interest. Databases are typically designed for understanding particular phenomena, to develop theoretical models, or to respond to specific needs for parameter regression. All participants emphasized the need to continue acquisition of property data necessary for chemical process design (such as thermochemical, thermophysical, transport properties, safety parameters, *etc.*).

Data acquisition and uncertainties

Two journal editors (from the Fluid Phase Equilibria and the Journal of Chemical and Engineering Data) participated in the panel. They highlighted two crucial observations from their editorial experience:

1. The quality of experimental data submitted for publication does not always meet the desired standard, and few experimentalists provide realistic and justified uncertainty information.

2. These problems have been recognized and at least partially mitigated by the collaboration between the National Institute of Standards and Technology (NIST) and five journals that publish thermodynamic and thermophysical property data. The editors emphasized that experimental papers in the journals have improved following the project's start in 2009 [14], and are expected to continue to improve as a result of this fruitful collaboration.

According to the panelist and audience opinion, the first observation can be attributed to several factors:

- Experimental procedure: Today, highly automated pieces of equipment are available on the market, resulting in operators relying on default features proposed by the vendors. As a result, their knowledge of the actual phenomena occurring in their equipment may be lost. This is often the case when the measurements are performed by students with insufficient input from expert advisors. While doing the measurements, it is very important to master all details of the experimental procedure, to understand the theory, to know the purity of the studied compounds and how the data are measured, *etc*. As an example, one panelist described a case of an absorption apparatus that was used to measure gas solubility in a polymer. When analyzing the data, the swelling of the solvent was not considered, which led to high inaccuracies in the results. A second example mentioned during the discussion concerned the measurement of phase equilibria with ionic liquids at high pressure (50 MPa). The researcher assumed that ionic liquids are absent from the vapour phase, as a result of neglecting the gas-phase non-ideality and the Poynting correction in the liquid fugacity calculation. A correct calculation would have shown that at such high pressures the vapour may contain up to 10 mol % of such compounds.

- Research traditions: Three tendencies affecting experimental practice in the scientific community were highlighted in the discussion – research traditions of high-quality measurements are frequently either lost, not yet gained, or compromised.

- In the past, most measurements were done with instruments constructed in house by experts and specifically designed to study specific systems and processes with careful consideration of possible complications and ways to avoid/compensate them. The above-mentioned widespread use of commercial instruments creates a dangerous illusion that everything can be easily studied with their use. In reality, such instruments can frequently handle only simple, typical measurements. Because of that, many sophisticated techniques developed for more complex cases are not used anymore, and the corresponding knowledge is lost. A relevant example is combustion calorimetry for organic compounds containing elements other than C, H, O, N: those measurements require a rotating-bomb calorimeter with additives (*e.g.*, oxidizing or reducing agents) and a thorough analysis of reaction products – a technique that is not widely commercialized and now exists only in a couple laboratories in the world. It is getting more and more common that commercial static-bomb calorimeters are used for those compounds, but this improper application results in unreliable data (ill-defined final state), and, unfortunately, the quality of such values is not questioned by researchers, reviewers, and/or data users.

- The second tendency is related to the international expansion of scientific research in general. Data growth is exponential, and this is connected to the fact that many newcomers appear in the field. Very often, they start from scratch without sufficient prior knowledge and established traditions on how to conduct research and how to report data. This leads to numerous manuscripts of low quality submitted to scientific journals, and, in some cases, publications of low quality being published.

- The nature of thermodynamic/thermophysical measurements stands behind the third tendency. This field is a traditional science, where sensational discoveries are not expected frequently. It is observed that in order to justify their activities, researchers in the field try to increase the number of measurements and make them simpler and cheaper, which sometimes compromises the data quality. In addition, a disproportionate amount of time is now spent by experts for grant application, rather than for research. As a result, they do not spend sufficient time and resources to analyze the measured property values, to prove the chemical identity, purity, and stability of studied substances, to validate instrumentation and experimental technique. Comparison of independent measurements (sometimes even within the same research group) frequently reveals unacceptably large inconsistencies in reported values. An additional comment from a discussion participant was that with the advent of numerical property databases, comprehensive reviews with analysis of experimental data (as for example [15,16]) are rarely consulted by researchers, which may also compromise the quality of experimental data, if an experiment is based on unreliable data, randomly taken from such databases (*e.g.*, for calibration or verification).

- Reporting of experimental uncertainties: The widespread opinion of both the panel and audience is that not enough attention is given to the evaluation of the experimental uncertainties. Many databases have only recently taken this into

account. The panelists pointed out the importance of incorporating all the causes that contribute to the combined expanded uncertainty of the measured value; this includes uncertainties related to the equipment, to the procedure, to the sensors, to the purity of the materials, and to the measured state variables. As an example, when measuring a vapour pressure, all uncertainties are often considered to be on pressure, while temperature or the sample purity may also give significant contributions to the uncertainty. The main issue is that many experimentalists do not know how to analyze uncertainties, especially if they do not understand the method they are using for the measurements. Simple assessment how well existing literature data can be reproduced is not a common practice.

How should one quantify the uncertainties? Is there any gold standard for this? In 1995, the International Organization for Standardization (ISO) in Geneva published the 'Guide to the expression of uncertainty in measurement' [17]. Before that, simple treatments were the norm. The Guide's recommendations were adopted with minor changes as the U.S Guide to the Expression of Uncertainty in Measurement (1997) [18] and were summarized in the Guidelines for the Evaluation and Expression of Uncertainty in NIST Measurements Results [19], which can be freely downloaded from the web site (http://physics.nist.gov/cuu/). In the document, the definitions of all quantities are given together with the identified terms that may contribute to the uncertainty of a measurement. In addition, NIST has launched an online tool (Uncertainty Machine at https://uncertainty.nist.gov/) to assist researchers in uncertainty assessment based on the recommendations. Despite all the recommendation efforts, it was noted during the discussion that many experimentalists are still not familiar with the current uncertainty terminology or even the uncertainty concept itself.

The above documents detail how to evaluate the type A uncertainty (largely based on repeatability), but provide only general notes about non-statistical assessments (type B uncertainty), which require significant testing, validation, and expertise. One of the participants from the audience noted that the lack of clear instructions probably explains why researchers do not go beyond the first part of the uncertainty analysis, and often confuse repeatability with uncertainty. Repeatability is not a sufficient uncertainty estimate in most cases. It is also alarming that experimentalists frequently make bare uncertainty statements (without any justification) or cite uncertainties claimed by instrument manufacturers without consideration of all factors causing errors, such as purity, conditions of measurements, and calibration issues. This frequently results in unrealistically low uncertainty estimates. It should be noted that trusting a misleading uncertainty may be more harmful than estimation of possible errors by subsequent users of the data, when uncertainty is not reported at all, since the users would tend to make a more conservative assessment.

It was indicated by an audience member that the most devastating problematic contributions to the uncertainty arise from ill-characterized substances and samples, their instability, failure to reach the claimed equilibrium, and flawed/inappropriate techniques. Those contributions may exceed all reasonable uncertainties, sometimes by orders of magnitude, and are often not reported, possibly, not recognized as problems by the researchers. Additional information is always needed to reveal such contributions. One of many examples of this kind is viscosity of bis(2-ethylhexyl) phosphate, whose literature experimental data span over one order of magnitude (two examples are [20,21]).

A single measurement may not give any indication of potential problems. A range of consistency checks and data reconciliation methods, which should be used by professionals on a regular basis, are discussed during the Round Table: For example, comparison of several results by different authors is the simplest way to reveal data problems; or, additional resources are predictions and theoretical methods, correlations, and thermodynamic consistency with other properties.

Consistency analysis

Internal consistency means that all the data of a given compound or mixture must agree within mutual uncertainties according to the fundamental thermodynamic principles. For pure compounds, these are for example the Clapeyron equation for the vapour pressure, or simply the fact that all phase equilibrium curves (and therefore phase properties) merge in the triple point. For mixture *PTxy* data, the Gibbs-Duhem relationship is a well-known consistency test [22][23]. Similarly, the end points of the mixture behaviour should correspond to the pure component properties [24]. Similar consistency methods for other types of data (liquid-liquid equilibrium – LLE, solid-liquid equilibrium – SLE) are proposed and reviewed by Kang et al. [25]. EOS can be useful to check for consistency among properties for both pure compounds and mixtures.

External consistency is established by comparing the data with those of similar systems. Some authors propose mapping the behaviour of compounds of the same chemical family to observe trends, and hence evaluate the quality of the data on whether or not they agree with the trend [26,27]. This analysis requires substantial expertise, because unexpected deviations from family trends may occur (some examples can be the smallest family members, odd-even effects, or steric effects in molecules).

Cross-consistency of the data by making comparisons with similar existing measurements for the same system should be applied more frequently. Doing so may lead to the question of how to distinguish good and bad data. This is a complicated issue when the data originate from different sources. It was proposed to use a correlation, *e.g.*, an EOS or an activity coefficient model that makes it possible to analyze deviations between the data and the calculated values, *e.g.*, as proposed by Van Ness [28,29]. One of the panelists proposed to use a density correlation/model for the validation of volumetric data for compressed liquids. If the data do not follow a normal distribution function, alternative approaches, *e.g.*, Mathematical Gnostics [30], robust statistics [31] , can be applied to identify the outliers.

Predictive and theoretical methods are often more conservative and robust against serious errors (provided that they are adequately validated and appropriate to the system and states under consideration), so they are useful tools for identifying experimental outliers. Some theoretical models are even comparable with state-of-the-art experiments (for example, statistical thermodynamics, quantum chemistry [32]) and can serve as a solid basis for validation. Most predictive methods are, however, empirical. They are shown to work adequately within the applicability (calibration) domain, but may have unpredictable extrapolation behaviour. Since they are based on experimental data, the quality of underlying data governs the reliability of such methods. Hence, their application requires extensive expertise in these methods.

What data are important?

After discussing data acquisition and assessment in general, the Round Table panel requested opinion of the process engineers in the audience, the knowledge of which properties are critical in the process design, *i.e.*, which properties should be determined as precisely as possible. The general feeling of the audience was that the critical properties (T_c, P_c, V_c) need to be accurate, since they are used in predictions of thermodynamic and physical properties and also as input parameters in cubic equations of state [33] and therefore in process and reservoir engineering, design and optimization. Among phase equilibria, VLE is the most frequently faced in the process design. Very often, VLE are obtained by measuring bubble pressures. Yet, it was recalled by the panel that the properties of utmost importance for separation purposes are generally partition coefficients [8].

Another question raised in front of the audience was about automation and “big data”. The opinion was mixed. The majority in the audience considered “big data” fashionable, but was concerned that expertise is excluded from such analyses, and automated results may contain large errors that would be obvious for an expert; at best, they can be used as a correlation tool to interpolate among existing data. At the same time, some audience members indicated that there are several successful “big data” stories in the property data field (e.g., selection of potential refrigerants from 184000 molecules [34]). It was suggested that the optimal way with respect to the time spent and the quality of results is an expert analysis combined with automatic consistency checks.

Conclusion on data

As a conclusion, the importance of critical data evaluation (method assessment, consistency checks, uncertainty evaluation) was stressed. This may require data redundancy: it is recommended to compare similar data obtained by different researchers or different labs, because human factors may strongly affect the results (if an experimental protocol is poorly designed), *i.e.*, quality of the data is very much associated with the person and human factor. If one person is measuring, higher level of confidence of the measured result is required. If two people are measuring on the same or different apparatuses, slightly different results can be obtained.

Several tools exist for data quality analysis, which could be better advertised or taught, according to the panel: the Gibbs-Duhem rule, EOS, and other internal consistency tests are essential, but other tools should not be ignored: for example, cross-consistency with other data sets of (a) the same system at other temperatures/pressures in temperature/pressure series (b) functionally related systems (e.g. homologous series, or homologous series of similar compounds with different functional groups).

- **For model developers: Model development**

Simulation tools are widely used for product and process design. These tools contain a wide range of thermodynamic models among which the end-users must select the most appropriate one. Depending on the available data, one may choose either correlative models or predictive models [35].

In correlative approaches, a large number of data may be used to fit empirical or semi-empirical thermodynamic models. One may use simple polynomials, but models based upon molecular thermodynamics, such as activity coefficient models or EOS, are often superior for some uses, according to the panel.

On the other hand, predictive models may require less data fitting, since their construction is based on physical concepts and previously validated data that allow them to remain valid over a wide range of physical conditions. Examples are models based on statistical thermodynamics as the SAFT [36,37], or *ab initio* thermodynamic models, like COSMO-RS [38], force-field [39] or quantum-based simulations. The development of new predictive models requires a good understanding of the controlling phenomena. In this case, an important issue consists in the selection of the most appropriate data that contain the relevant information. Combining data of different origin (*e.g.*, different properties on the same system – mixing properties, or including molecular information) may allow an improved understanding of the true governing phenomena. A close collaboration between experimental labs and model developers may help improve such predictive modeling.

Parameter regression

Once a thermodynamic model has been chosen, its parameters must be determined. As an example, binary interaction parameters are mentioned: They are commonly used in thermodynamic models (EOS, activity coefficient models) to enable the prediction of phase equilibrium states, especially for non-ideal fluid systems [4]. Parameters regression is even more abundant for correlative approaches, which typically interpolate well but are not reliable in extrapolation.

The quality of the model depends to a great extent on parameter determination. Typical questions that must be addressed [40] are (1) what parameters and how many must be regressed (as opposed to a fixed value that can be determined on physical grounds, *e.g.*, σ-profile in COSMO-RS); (2) how the objective function should be constructed, and (3) what data should be used to validate and test the resulting model (validation and test sets).

Focusing on the question of the objective function to be used in this procedure, two issues were raised at the Round Table: (1) how to use the experimental uncertainties and (2) what property/data range is most significant in this procedure.

(1) It is very important to determine all the uncertainties (the errors) in parameter fitting, which is frequently a complicated task without a well-defined solution. This analysis yields uncertainties on the model parameters. These uncertainties then propagate through the process simulator to the process design at the end. An estimation of the model parameters and uncertainties associated with them can be performed by using the weighted least-square method with the minimization of an objective function based on the model, model parameters, experimental data and their uncertainties [33]. Linear or non-linear regression can be applied; according to the panel, the latter with a general non-linear form of each model for each property delivers better results. Monte Carlo technique can be further used to quantify uncertainties of the thermodynamic model by propagation of all input uncertainties, assuming that the distribution of the experimental data is known. Thermodynamic model uncertainties can then serve as input uncertainties to estimate uncertainties in the process simulation.

The panel concluded that the uncertainty assessment method described above is very complex and it is not very practical for industrial applications. A new approach to uncertainty analysis in process simulation, based on the Margules Uncertainty Analysis method was proposed in [8], adding perturbations to, *e.g.*, activity coefficients of components or other correlations. The method gives an insight into the influence of the uncertainty of the properties and their effect on the process design, demonstrated by two industrial examples provided in Ref. [6].

(2) Yet, the question of which property/data range has the most effect on the process simulation results is often overlooked. One example concerning the range of the data considered was mentioned: the quality of the model for high purity applications may be significantly different depending on whether or not infinite dilution properties are included in the regression. Similarly, it is well known that the regression results depend on which property is considered with a higher weight (e.g., bubble pressures or solubilities). Obviously, the key property from the process design point of view may depend on the specifics of the process considered. For example, it was noted during the discussion that phase appearance problems typically have higher sensitivity to binary parameters of extreme compounds (very heavy and very light in the case of VLE), while separation problems require an accurate description of relative volatilities of compounds close to the column cut-point.

Finally, it may be useful to consider using covariance analysis to avoid overlooking parameter coupling or incompleteness of data.

A story of one of the panelists seems worth bringing up as an illustration of the need to use proper weighting of experimental data as well as additional data (often including heats of mixing data [41] is extremely valuable when temperature-dependence is key) in the design of chemical plants [42]:

“I had isothermal VLE data at two temperatures, fitted temperature-dependent parameters to them in order to extrapolate to high temperatures and pressures for a preliminary design of their [customer-requested] process. The fit seemed fine, as well as the residuals. The data seemed also fine (they were measured in this group which I am now leading...). During extrapolation, however, I realized that the temperature dependency of the azeotrope was initially even qualitatively wrong, leading to completely erroneous design of the distillation column (separating ethanol and light hydrocarbons). After looking carefully at the points near the azeotropic point, I got the problem resolved by changing the weighting of the points in the parameter fitting. We did not have enthalpy data then, or VLE at very wide temperature ranges. I think this is a case where it is not really about error propagation, but something closely related.”

- **For vendors of simulation tools**

Companies developing process simulation software provide an efficient and convenient tool to practicing engineers so as to help them reduce cost and improve their design. Process simulators typically contain models and parameter databanks that have been fitted to available data. Often, the companies do not sell data, but rather an engine that is expected to be used in a responsible way. The panelists drew an analogy with car manufacturers, who are typically not responsible if a driver drinks and drives: The vendor companies are

expected to not be liable if their tool is used in an irresponsible way. The prerequisite for each simulation package to be used effectively is a certain level of knowledge, which starts with the basics of thermodynamics and chemical engineering.

It is clear that some property calculations may be wrong if they are extrapolated too far from the fitted data. A particular challenge is represented by the temperature-dependent properties (such as vapour pressures), where a small uncertainty in a given temperature range, where the experimental data are available, may lead to a very high uncertainty, for example, upon extrapolation to the triple point (even 100 % uncertainty). The various companies that were represented in the discussion had different points of view on their responsibility regarding the information to be transferred to the user. Often, they rely on other providers as for example the widely accepted correlations proposed by the DIPPR consortium [43]. Some suggested indicating the uncertainties on critical point parameters, since they are crucial for a number of calculated thermodynamic and physical properties: as indicated by them, small uncertainties in critical point properties that are used in EOS may have a dramatic effect on the final calculated results. The idea to include uncertainty estimations as an output of the calculation is however not often implemented in current process simulation software.

An opinion from the audience was that in some cases it may be useful that a warning be available when the calculation range exceeds the experimental range. This solution was dismissed as being ineffective.

The panel suggested that the most effective approach at this stage would be to leave the responsibility to users to perform their own regression by providing adequate tools and experimental data.

- **For process engineers as designers**

The panel suggested that the process engineer who uses the process simulator for design or improvement of a process is the one who bears the responsibility for the quality of the design. He or she is accountable for the results. In the end, a correct analysis of the impact of data on the final result is a question of how much risk one is willing to take. This requires correct assessment of the quality of data used as an input.

Yet, unfortunately, it appears to several panelists and discussion participants that the users of simulation packages may not be aware of the quality of data they use, the model limitations, the origin of the parameters, *etc*. For some users, the process simulator is a black box. This brings up the issue of chemical engineering education: Several university professors noted that it is increasingly challenging to teach expertise when the trend of the upcoming generation is to rely on the results provided by a computer. The panel recommended that teachers and their host institutions never abdicate their responsibilities, and that they encourage their students to recognize the importance of data in all their process calculations.

The general opinion of the panel and audience is that the process simulation should always start with the data and understanding of their origin. The user may benefit from knowledge of how the data were obtained. Process engineers would typically not be in a position to

measure and analyze all of the property data that go into a process design. But, the panel concluded that in the end, they need to be very careful and responsible, since their company could lose a lot of money or cause large damage through a wrong decision based on incorrect use of a process simulator. Some discussion participants suggested that in case of any doubts about the reliability of the data to be used, the data analysis performed, or any other aspects of the simulation, the process engineers should seek help from known data experts in the field or data evaluation professionals, whose job is to provide recommended property values. This additional time and resource investment could pay off and reduce catastrophic consequences.

An example of wrong decision-making is the explosion in Texas that was described above: The properties of the mixture were not known, and the designers were not aware of the fact that there was a lack of data on the system at hand.

Paths for improvements

- **Connections between different levels**

An important conclusion from the discussion relates to the large gap in understanding of the need of the actors at the different levels of the development described above: The process engineer may have little understanding of the way data are measured, and the experimentalist may have never seen a distillation column. It is natural that expertise of everyone is limited, but it appears essential to know one's own limitations and be curious about the culture of the other contributors to the process.

The panel was aware of only a few investigations that have been published showing the impact of data uncertainty, through the parameter uncertainties, all the way to the uncertainties on the process design [3,4,13,33]. Such investigations may help enormously in decision making.

Better communication among different levels (experimentalists, data evaluation professionals, model developers, simulation companies, and process engineers) was proposed to improve data quality and assessment, validation methods, and error propagation through the simulation software.

- **Teaching**

According to the panel, education in thermodynamics and in simulation plays a key role in understanding the impact of property data uncertainties on process development and plant operations. Based on the panel and audience opinion, this is a global issue. One university professor from the audience mentioned that in the UK, PhD students in chemical engineering are taught about the source of data and how to use them. It was also suggested not to hesitate to go back to the literature to allow a critical analysis.

Students need to understand properly the models and algorithms that underpin a process simulator and be aware of the limitations of such simulators. Not only the best cases, but also the worst cases should be discussed. All engineers must learn to be responsible, as part

of the typical codes of ethics. The panelists believe that perhaps an important point to teach is modesty: know your limitations and do not hesitate to go talk with the experts at any other level in the chain of process development.

Figure 1 illustrates the key messages that can be retained from the discussion: each level of stakeholders has its own quality indicators, but in order to reach a successful process design, effective communication and learning from each other is needed.

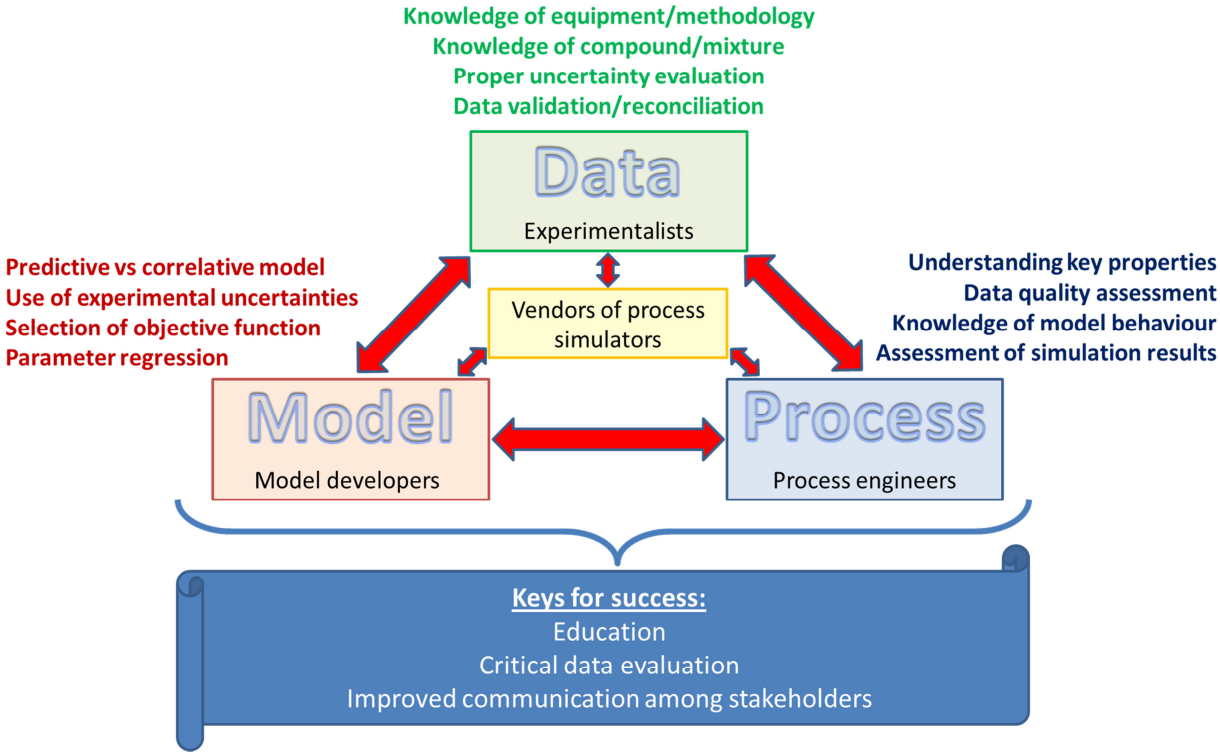

Figure 1: the main outcomes of the round-table discussion

Conclusion

The Round Table discussion raised several important messages which may be summarized as follows:

- Data reconciliation methods should be used, such as consistency analysis, use of predictive methods, *etc*., to assess the quality of data which enter a process simulator. Regarding the physical properties, panelists suggested that the biggest uncertainty is often not connected with the fitting parameters, but rather with the original property data. It was also suggested that vendors should provide uncertainty estimation software and sensitivity analysis to be built in the simulation packages.

- Communication across traditional boundaries should be facilitated to remove barriers towards increased implementation of the uncertainty concept, from the data measured up to the process design.

- Teaching is a key word: Any user of process simulation software should have a good understanding of the thermodynamic fundamentals and the concept of uncertainty.

Acknowledgements:

The authors, organizers, and chair of the Round Table discussion, want to thank the panelists for their contributions: Jens Abildskov (DTU Chemical Engineering, Lyngby, Denmark), Ville Alopaeus (Aalto University, Aalto, Finland), Frank Eckert (COSMOlogic GmbH & Co, Leverkusen, Germany), Cor Peters (Khalifa University of Science and Technology, Abu Dhabi, UAE), Zdenek Wagner (E. Hála Laboratory of Separation Processes, Suchdol, Czech Republic). Special thanks also to all the participants to the discussion, since many useful thoughts came from the audience.

This article is a partial contribution of the National Institute of Standards and Technology (NIST) and is not subject to copyright in the United States for the author A.B. Trade names are provided only to specify procedures adequately and do not imply endorsement by the National Institute of Standards and Technology. Similar products by other manufacturers may be found to work as well or better.

References

[1] S. Kim, J.W. Kang, K. Kroenlein, J.W. Magee, V. Dicky, D. Frenkel, Chem. Eng. Educ. 47 (2013) 48–57.
[2] J.W. Magee, S.H. Kim, J.W. Kang, K. Kroenlein, V. Diky, C.D. Muzny, A.F. Kazakov, R.D. Chirico, M. Frenkel, 10th European Congress of Chemical Engineering (2015, Nice, France).
[3] J. Frutiger, I. Bell, J.P. O'Connell, K. Kroenlein, J. Abildskov, G. Sin, Molecular Physics 115 (2017) 1225–1244.
[4] S. Hajipour, M.A. Satyro, M.W. Foley, Fluid Phase Equilibria 364 (2014) 15–30.
[5] S. Hajipour, M.A. Satyro, M.W. Foley, Energy Fuels 28 (2014) 1569–1578.
[6] S. Watanasiri, Pure and Applied Chemistry 83, 6 (2011) 1255–1281.
[7] V. Diky, R.D. Chirico, C.D. Muzny, A.F. Kazakov, K. Kroenlein, J.W. Magee, I. Abdulagatov, J.W. Kang, R. Gani, M. Frenkel, J. Chem. Inf. Model. 53 (2013) 249–266.
[8] P.M. Mathias, J. Chem. Eng. Data 59 (2014) 1006–1015.
[9] H.C. Jarvis, Chem. Eng. Progress 67 (1971) 41–44.
[10] R.H. Freeman, M.P. McCready, Chem. Eng. Progress 67 (1971) 45–50.
[11] R.W. King (Ed.), Safety in the Process Industry, Butterworth-Heinemann Ltd., London, 1990.
[12] C.J. Peters, Personal communication, Barcelona.
[13] J. Frutiger, J. Andreasen, W. Liu, H. Spliethoff, F. Haglind, J. Abildskov, G. Sin, Energy 109 (2016) 987–997.
[14] P.T. Cummings, Fluid Phase Equilibria 276 (2009) 165–166.
[15] R. Dohrn, J.M. Fonseca, S. Peper, Annu Review Chem Biomol 3 (2012) 343–367.
[16] J.M. Fonseca, R. Dohrn, S. Peper, Fluid Phase Equilibria 300 (2011) 1–69.
[17] JCGM 100:2008, Evaluation of measurement data — Guide to the expression of uncertainty in measurement. https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf.

[18] NCLS International, U.S. Guide to the expression of uncertainty in measurement, Boulder, Colorado.
[19] B.N. Taylor, C.E. Kuyatt, Guidelines for the Evaluation and Expression of Uncertainty in NIST Measurement Results, Gaithersburg, MD, 1994.
[20] L.R. Koekemoer, M.J.G. Badenhorst, R.C. Everson, J. Chem. Eng. Data 50 (2005) 587–590.
[21] N. Swain, S.K. Singh, D. Panda, V. Chakravortty, Journal of Molecular Liquids 85 (2000) 321–330.
[22] J.M. Prausnitz, R.N. Lichtenthaler, E. Gomes de Azevedo, Molecular Thermodynamics of Fluid Phase Equilibria, Prentice Hall Int, 1999.
[23] H.C. van Ness, Pure and Applied Chemistry 67 (1995) 859–872.
[24] J.W. Kang, V. Diky, R.D. Chirico, J.W. Magee, C.D. Muzny, I. Abdulagatov, A.F. Kazakov, M. Frenkel, J. Chem. Eng. Data 55 (2010) 3631–3640.
[25] J.W. Kang, V. Diky, R.D. Chirico, J.W. Magee, C.D. Muzny, A.F. Kazakov, K. Kroenlein, M. Frenkel, J. Chem. Eng. Data 59 (2014) 2283–2293.
[26] J. Rozmus, J.C. de Hemptinne, N. Ferrando, P. Mougin, Fluid Phase Equilibria 329 (2012) 78–85.
[27] P.M. Mathias, S. Watanasiri, J. Chem. Eng. Data 56 (2011) 1658–1665.
[28] H.C. van Ness, S.M. Byer, R.E. Gibbs, AIChE J. 19 (1973) 238–244.
[29] P.L. Jackson, R.A. Wilsak, Fluid Phase Equilibria 103 (1995) 155–197.
[30] Z. Wagner, M. Bendová, J. Rotrekl, P. Velíšek, J. Storch, P. Uchytil, K. Setnickova, J. Řezníčková, Journal of Solution Chemistry 46 (2017) 1836–1853.
[31] A. Gelman, J. Carlin, H.S. Stern, D. Dunson, A. Vehtari, D. Rubin, Bayesian Data Analysis, 3rd ed., CRC Press, 2013.
[32] E. Paulechka, A. Kazakov, J. Phys. Chem. A 121 (2017) 4379–4387.
[33] S. Hajipour, M.A. Satyro, Fluid Phase Equilibria 307 (2011) 78–94.
[34] M.O. McLinden, J.S. Brown, R. Brignoli, A.F. Kazakov, P.A. Domanski, Nature Communications 8 (2017) 14476.
[35] J.-C. de Hemptinne, J.-M. Ledanois, P. Mougin, A. Barreau, Select Thermodynamic Models for Process Simulation - A Practical Guide using a Three Steps Methodology, Technip, 2012.
[36] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, Fluid Phase Equilibria 52 (1989) 31–38.
[37] G. Jackson, W.G. Chapman, K.E. Gubbins, Mol.Phys. 65 (1988) 1–31.
[38] A. Klamt, COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design, Elsevier Science & Technology, Amsterdam, 2005.
[39] P. Ungerer, B. Tavitian, A. Boutin, Applications of Molecular Simulation in the Oil and Gas Industry, Editions Technip, 2005.
[40] P. Englezos, N. Kalogerakis, Applied Parameter Estimation for Chemical Engineers, Marcel Dekker,Inc, New York, Basel, 2001.
[41] H. Renon, L. Asselineau, G. Cohen, C. Raimbault, Calcul sur Ordinateur des équilibres liquide-vapeur et liquide-liquide, Publication de l'Institut Français du pétrole, Paris, 1971.
[42] V. Alopaeous, Personal communication, Barcelona.
[43] J.R. Rowley, V. Wilding, J.L. Oscarson, Y. Yang, N.F. Giles, DIPPR (R) Data Compilation of Pure Chemical Properties.