Automatized determination of fundamental equations of state based on molecular simulations in the cloud

Andreas Köster[a], Tao Jiang[b], Gábor Rutkai[a], Colin W. Glass[b], Jadran Vrabec[a,*]

[a]*Lehrstuhl für Thermodynamik und Energietechnik, Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany*
[b]*Höchstleistungsrechenzentrum, Universität Stuttgart (HLRS), 70550 Stuttgart, Germany*

Abstract

An automatized procedure for the parameterization of fundamental equations of state (EOS) that are explicit in terms of the Helmholtz energy and are based on molecular simulation data is presented. The simulation runs are carried out via a cloud-based framework that combines multiple, distributed computing resources. A user-friendly graphical user interface ensures that minimal knowledge about the background operations is required. In order to exemplify the capabilities of this approach an EOS for ethylene oxide is created and compared to data from the literature.

Keywords: thermodynamic properties, fundamental equations of state, molecular modelling and simulation, ethylene oxide, cloud computing

1. Introduction

Thermodynamic data are essential for the design and optimization of chemical engineering processes. Typically, the chemical industry is relying on experimental investigations to generate such data. Depending on the type of fluid, this undertaking is difficult to conduct or in the case of hazardous substances, e.g. being explosive, toxic or mutagenic, this might even be practically impossible. As an attractive alternative route, molecular modelling and simulation on the basis of classical force fields has been proven to be an adequate method due to the steady developments of the last decades. Along with its powerful predictive capabilities, large datasets of thermophysical data can be rapidly produced with considerably less effort than laboratory measurements.

It has recently been shown that molecular simulation data are useful for the construction of empirical fundamental equation of state (EOS) correlations that are explicit in terms of a thermodynamic potential [1, 2, 3, 4], which is most often the Helmholtz energy [5]. Thermodynamic potentials are essential building blocks of thermodynamics, because every other time-independent thermodynamic property can be obtained as a combination of their partial derivatives with respect to their independent variables. Unfortunately, the exact mathematical expression of a thermodynamic potential, with the exception of very simple systems (e.g. ideal gas or some hard-body fluids), is unknown. The construction of an EOS therefore necessarily means that the mathematical form of such a correlation has to be fitted to thermodynamic data. Once the parameterized correlation is available, it

*Corresponding author: Jadran Vrabec, Warburger Str. 100, 33098 Paderborn, Germany, Tel.: +49-5251/60-2421, Fax: +49-5251/60-3522, Email: jadran.vrabec@upb.de

provides thermodynamic data in a consistent manner and can be used to inter- and extrapolate between or beyond the underlying data points. More importantly, EOS have the ability to yield all thermodynamic data, and not just the ones considered in their fitting process. Naturally, the more thermodynamically independent properties are considered during the construction of EOS, the better the representation of other properties. Therefore, in order to create such a correlation, a large number of independent thermodynamic data points has to be available. This scenario is the application field for the statistical mechanical formalism proposed by Lustig [6, 7]. With this formalism, the derivatives of the Helmholtz energy up to arbitrary order can be obtained at a given state point by a single canonical (NVT) ensemble molecular simulation run [6, 7]. This formalism also exists for the entropy derivatives in the microcanonical (NVE) ensemble [6, 7]. Since any thermodynamic property is just a combination of these derivatives, this represents the full set of thermodynamically independent data. Due to this improvement on availability, reliable EOS can very efficiently be fitted on the basis of molecular simulation data only.

However, setting up molecular simulation runs and fitting EOS requires considerable manual effort and expertise. Maginn [8] pointed out that molecular simulation has to overcome some barriers in order to become a mainstream tool in chemical engineering. One of these barriers is the time which is necessary to set up and analyze a molecular simulation run compared to its actual runtime. Furthermore, there are few software packages which are able to yield all properties that are important for the chemical industry (e.g. phase equilibria, gas solubility, heat capacity, density, etc.) [8]. Hence, the goal of this work is to automatize and optimize this workflow by combining the molecular simulation tool $ms2$ [9] with an EOS fitting algorithm and embedding this into a cloud-based simulation framework which requires minimal knowledge about the background operations, and has a simple graphical user interface (GUI) for creating the necessary input for simulations. In the background, a job manager efficiently allocates the molecular simulation runs to the available high performance computing (HPC) resources. Once the simulation runs are terminated and the results are ready, the empirical EOS correlation is fitted and its parameters are returned, along with the quality of representation of the simulation data. Thus, the development of EOS is essentially becoming an automatic process. In this work, the capabilities of this approach and the resulting EOS are compared for ethylene oxide with experimental data from the literature and with the recently published EOS by Thol et al. [4] for this fluid, that we consider as a reference here.

2. Fundamental equation of state correlation

The present EOS correlation is explicit in terms of the Helmholtz energy a. It is common to separate a into an ideal contribution a^{o} and a residual contribution a^{r}

$$\alpha(T,\rho) = \frac{a^{\mathrm{o}}(T,\rho) + a^{\mathrm{r}}(T,\rho)}{RT} = \alpha^{\mathrm{o}}(T,\rho) + \alpha^{\mathrm{r}}(T,\rho) = \alpha^{\mathrm{o}}(\tau,\delta) + \alpha^{\mathrm{r}}(\tau,\delta), \tag{1}$$

with the inverse reduced temperature $\tau = T_{\mathrm{c}}/T$, the reduced density $\delta = \rho/\rho_{\mathrm{c}}$ and the molar gas constant $R = 8.3144621\,\mathrm{J}\cdot\mathrm{mol}^{-1}\cdot\mathrm{K}^{-1}$ [10], where T_{c} is the critical temperature and ρ_{c} is the critical density. The ideal contribution a^{o}, which only depends on the properties of single molecules, can usually be obtained from

spectroscopic data or ab initio calculations, whereas the much more demanding residual contribution is the target of molecular simulation because it depends on the interactions between the molecules. The empirical equation that represents the residual contribution to the reduced Helmholtz energy α^{r} consists of polynomial and exponential terms

$$\alpha^{\mathrm{r}}(\tau,\delta) = \sum_{k=1}^{N_{\mathrm{pol}}} n_k \tau^{t_k} \delta^{d_k} + \sum_{k=N_{\mathrm{pol}+1}}^{N_{\mathrm{exp}}} n_k \tau^{t_k} \delta^{d_k} \exp(-\delta^{l_k}). \tag{2}$$

In general, the simultaneous optimization of coefficients n_k, exponents t_k, d_k, l_k as well as the number and type of polynomial and exponential terms requires expert use of complex non-linear fitting algorithms [5, 11]. Assuming that a sufficient amount of thermodynamic data is available, such an effort usually takes months or years, requiring a high level of human intervention and experience. The recently published EOS of Thol et al. [4] for ethylene oxide is such a model, and was therefore chosen as a reference in the present work. In the fitting process of that EOS [4], both experimental and molecular simulation data were used. Nonetheless, such an approach is currently not feasible in an automatized EOS development process. There are, however, recommendations in the literature for the functional form and the values of the exponents t_k, d_k, l_k of correlations that have proven to be a good choice for a large number of fluids. These are for example the 10- and 12-term EOS by Span and Wagner [5, 12], the 14-term EOS by Sun and Ely [13] and the MBWR EOS [14], which consists of 40 terms when transformed into the Helmholtz energy representation [5]. In this work the MBWR EOS [14] was chosen to represent the molecular simulation data set. Such generalized EOS have the advantage that, since the exponents t_k, d_k, l_k in Eq. (2) are known, the task at hand reduces to the fitting of the coefficients n_k, which can be done in a fully automatized manner and has a response time of seconds, e.g., with the algorithm for weighted multiproperty fits of Hust and McCarty [15], in which the weight corresponds to the statistical uncertainty of the underlying molecular simulation data. The values of the coefficients t_k, d_k, l_k and n_k for the present EOS can be found in table 1.

3. Software Architecture

Due to the volume of calculations in the field of molecular simulation, researchers typically rely on large computing infrastructure. Here, a single molecular simulation run is relatively time intensive and the database required to parameterize an EOS should contain results from a large number of independent simulation runs at different state points that cover a large part of the fluid region. Thus, the total computational effort is extensive. However, besides the scalability of an individual simulation run, multiple instances for the different state points can be executed in an embarrassingly parallel fashion. Therefore, the simulation can be scaled both in horizontal and vertical direction, i.e. with respect to number of instances and parallel execution of a single instance.

A highly accurate calculation of thermodynamic properties at a given state point with *ms*2 currently requires approximately 20 hours on 32 cores. For one EOS, between 100 and 1'000 state points need to be evaluated. Thus, even with exclusive access to a small compute cluster, the calculation can take months. At the same

Table 1: Parameters for the residual part of the reduced Helmholtz energy $\alpha^{\mathrm{r}}(\tau, \delta)$ according to Eq. (2). Note that only the coefficients n_k were fitted in the automatized procedure, the exponents t_k, d_k, l_k were taken from the literature [14].

k	n_k	t_k	d_k	l_k
1	-0.444761941	0	1	0
2	4.295500322	0.5	1	0
3	-6.304073590	1	1	0
4	2.495663889	2	1	0
5	-1.413871625	3	1	0
6	0.122634301	0	2	0
7	-0.403993247	1	2	0
8	0.953802102	2	2	0
9	3.626258062	3	2	0
10	-0.005286229	0	3	0
11	0.167724139	1	3	0
12	-0.261302669	2	3	0
13	0.037473360	1	4	0
14	-0.148085338	2	5	0
15	-0.112161698	3	5	0
16	0.056427170	2	6	0
17	-0.005390223	2	7	0
18	0.010228306	3	7	0
19	-0.001370416	3	8	0
20	-15.487760440	3	0	0
21	-0.079331908	4	0	0
22	0.006185345	5	0	0
23	15.486057470	3	0	2
24	0.081647538	4	0	2
25	-0.007000321	5	0	2
26	10.769184560	3	2	2
27	1.108403574	4	2	2
28	-0.210854465	5	2	2
29	3.440671002	3	4	2
30	0.841327903	4	4	2
31	-0.256828769	5	4	2
32	0.416506137	3	6	2
33	0.492487787	4	6	2
34	-0.007594133	5	6	2
35	0.123791118	3	8	2
36	0.027697626	4	8	2
37	-0.042799502	5	8	2
38	-0.005719867	3	10	2
39	0.022433309	4	10	2
40	0.002604667	5	10	2

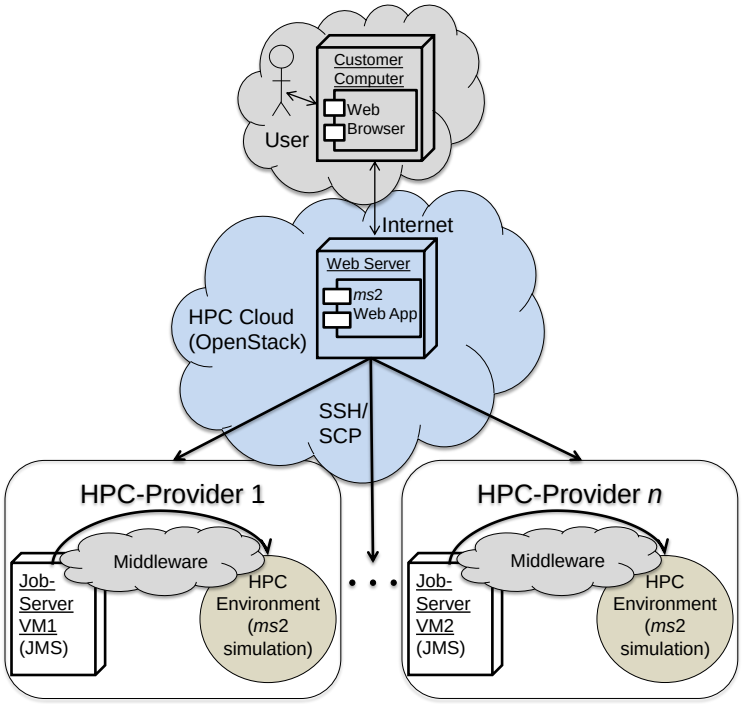

Figure 1: Distributed execution of molecular simulation runs through the High Performance Computing (HPC) cloud. SSH: Secure Shell; SCP: Secure Copy Protocol; JMS: Java Messaging Service.

time, filling a HPC machine with relatively small computing jobs is not viable, because these machines are intended for large scale jobs. Typically, a limited number of small jobs is welcomed, as this facilitates charging HPC machines to capacity. Therefore, distributing the work load across multiple machines is the most sensible approach, cf Figure 1.

To enable the execution of such a large scale simulation workflow, a significant amount of heterogeneous compute resources needs to be allocated dynamically. Here, the simulation workflow was concurrently deployed across multiple HPC systems on a cloud infrastructure. This cloud-based molecular simulation framework is aware of the available resources, initiates the simulations automatically and adapts dynamically to the actual execution time of individual simulation runs. A web browser interface allows for straightforward deployment from a regular workstation. The user is guided through the simulation setup with explanations, requiring minimal expertise.

Besides integrating the simulation software $ms2$, the framework introduces an additional user-controlled cloud environment, offering personalized and isolated front-ends with more flexibility in comparison to traditional HPC login nodes. Custom images can be deployed, root access be granted and additional services (e.g. MySQL, license server) as well as infrastructures (virtual machine internal networks, persistent/shared storage, external firewalls, etc.) can be operated directly. Another advantage is that users can submit their jobs to different HPC systems through a single application programming interface (API), since the ff-Middleware layer provides an abstraction of the actual batch-system. The present system architecture introduces a three-layer structure, i.e. web user interface, web application/service and calculation layer, cf. Figure 2.

Web User Interface: The web-based GUI runs as a part of the web browser on e.g. a desktop computer using

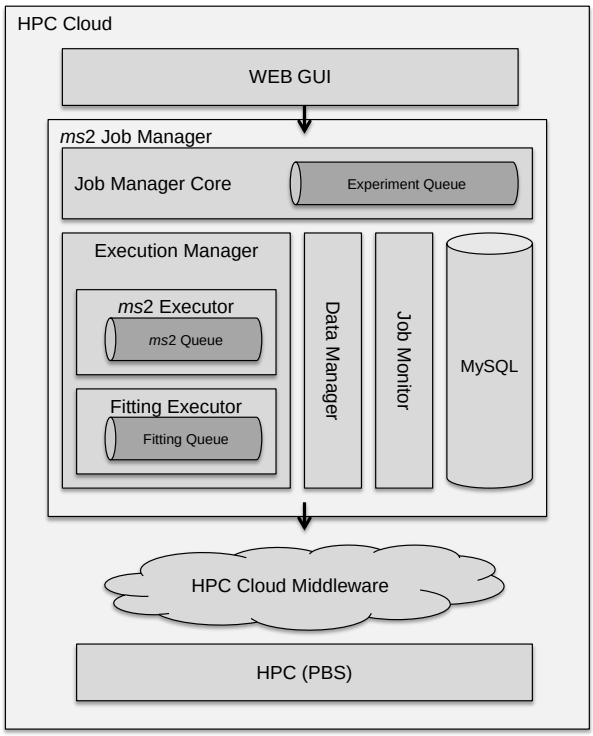

Figure 2: Components of the *ms*2 Job Manager.

advanced HTML5 and JavaScript techniques in order to provide a dynamic user interface, which allows users to configure, execute, monitor and view the results of the molecular simulation instances running in the cloud.

Job Manager: On the middle layer, a software component called $ms2$ Job Manager was designed to provide a set of services, implementing a Service Oriented Architecture (SOA), which is exploited both from internal and external components and allows for better scalability, maintainability and availability. The provided services allow to:

- Prepare and upload/download input/output files to/from the HPC systems in the background (Data Manager)
- Submit the simulation jobs to the portable batch system (PBS) queue on HPC systems (Execution Manager)
- Monitor the job execution (Job Monitor)
- Store the history data in terms of simulation configurations, input/output data, provenance and tracking information (MySQL database)

Thus, users can perform complex simulation workflows on distributed computing resources by simply clicking a few buttons. Therefore, many of the obstacles that currently confound the delivery, accessibility and usability of traditional, non-service-oriented simulation applications are avoided.

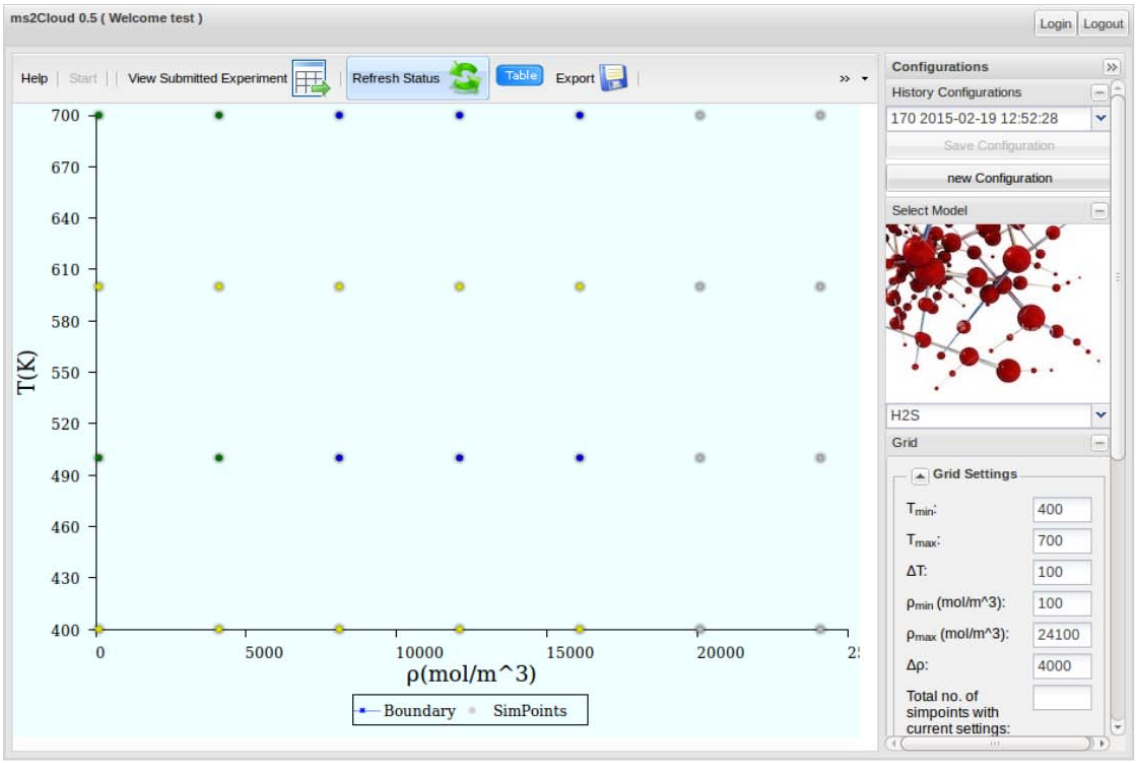

Figure 3: Web interface of the cloud-enabled molecular simulation. First step: specification of simulation conditions.

4. Workflow

The procedure of the cloud-based EOS creation is briefly described in the following. In a first step, the molecular force field model and the state points are specified, cf. Figure 3. This can be done by defining a grid or by clicking into the temperature-density plane. It is important to distribute the state points relatively evenly, covering the entire homogeneous fluid region of interest. In a preceding publication [2], it was shown that at least 100 state points should be specified for that purpose. Although vapor-liquid equilibrium (VLE) data are generally employed in EOS fitting, they are not necessarily required here if the homogeneous region is very well sampled. Additionally, some simulation settings, like the number of simulation steps and the number of compute nodes, have to be supplied. When the start button is activated, the calculations are distributed to the available resources and the status of each simulation is indicated by the color of the particular state point (grey: pending, yellow: queued, blue: running, green: successfully terminated).

When the simulations are completed, the fitting GUI can be opened, cf. Figure 4. In addition to some basic options, like type of EOS or number of iterations, some fluid specific settings, e.g. a rough estimate of the fluid's critical point for EOS fitting, have to be specified. If the critical point is not known, it is important to initially start fitting with an isotherm which is likely to be higher than the critical temperature. Otherwise, the isotherm could contain state points in the vapor-liquid two-phase region.

As a result of this procedure, the saturated liquid and saturated vapor density curves are displayed, cf. Figure 5. It is possible to restart the fitting with adjusted parameters or return to the previous GUI to add further state points.

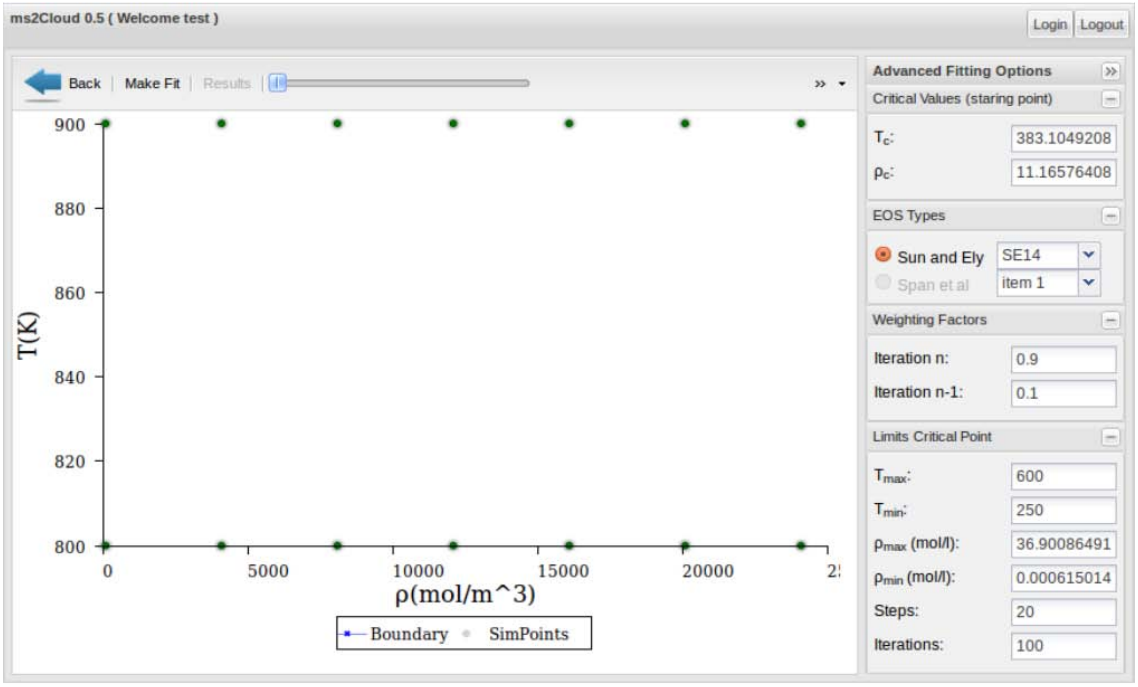

Figure 4: Web interface of the cloud-enabled molecular simulation. Second step: initial settings for EOS fitting.

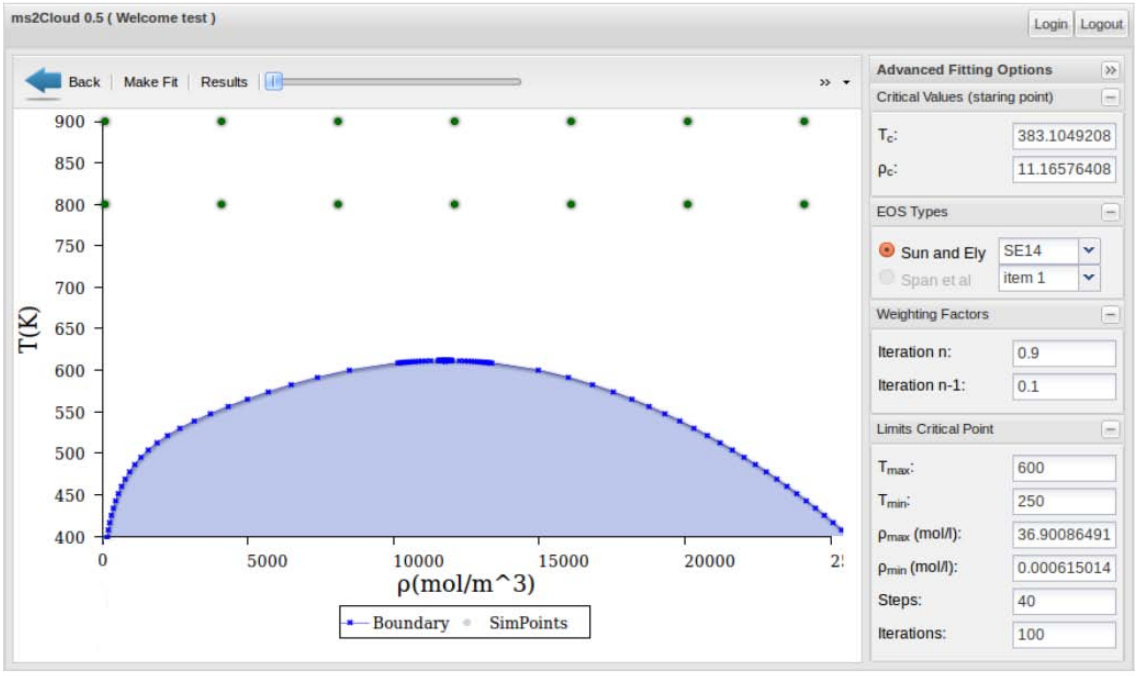

Figure 5: Web interface of the cloud-enabled molecular simulation. Third step: adjustment of the fitting details.

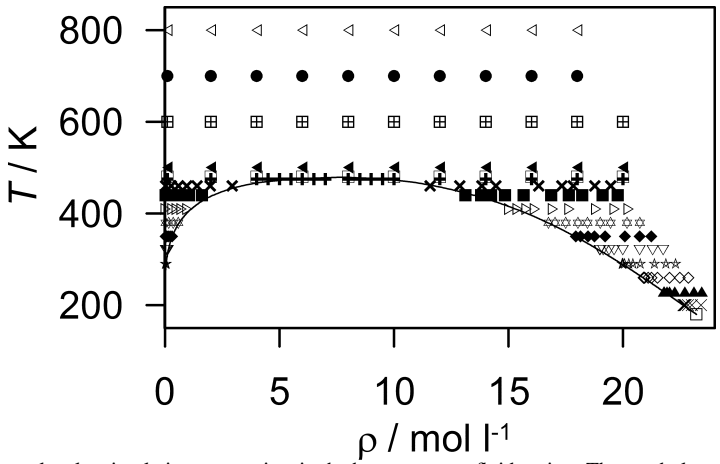

Figure 6: Distribution of the molecular simulation state points in the homogeneous fluid region. The symbols represent different isotherms.

5. Underlying molecular simulation data

The automatized creation of the simulation data set is based on the statistical mechanical formalism proposed by Lustig [6, 7] which was used here in its NVT ensemble form. This formalism was designed to yield any residual Helmholtz energy derivative A^{r}_{mn} (for $m > 0$ or $n > 0$) at a given state point, specified by its temperature and density, from a single simulation run per state point. These derivatives are exactly the properties that can be obtained by deriving eq. (2) with respect to its independent variables, i.e. reduced inverse temperature τ and reduced density δ

$$A^{\mathrm{r}}_{mn} = \tau^m \delta^n \frac{\partial^{m+n}\alpha^{\mathrm{r}}(\tau,\delta)}{\partial\tau^m \partial\delta^n} = (1/T)^m \rho^n \frac{\partial^{m+n}\alpha^{\mathrm{r}}(1/T,\rho)}{\partial(1/T)^m \partial\rho^n}, \tag{3}$$

because in $\tau = T_{\mathrm{c}}/T$ and in $\delta = \rho/\rho_{\mathrm{c}}$ the critical temperature T_{c} and the critical density ρ_{c} are constants. Consequently, each A^{r}_{mn} result from a molecular simulation run represents a new analytical derivative of eq. (2) and stands for a new equation. The complete set of such equations forms a system of linear equations that has to be solved for the coefficients n_k, which can be done with the algorithm of Hust and McCarty [15].

The dataset of the present work was generated using the molecular simulation tool $ms2$ [9]. For each sampled state point $ms2$ yields every of the following derivatives: A^{r}_{10}, A^{r}_{01}, A^{r}_{20}, A^{r}_{11}, A^{r}_{02}, A^{r}_{30}, A^{r}_{21}, A^{r}_{12}. However, it is not always necessary to include all of them into the EOS fitting in order to have a good representation for all of them. Here, only four derivatives A^{r}_{10}, A^{r}_{01}, A^{r}_{20}, A^{r}_{11} at 205 state points in the homogeneous fluid region were used, cf. Figure 6. Since all of these derivatives are thermodynamically independent, the database for the EOS fitting process is large (820 independent data points).

A single NVT Monte Carlo simulation [16] was performed at each state point and then sampled according to Lustig's formalism [6, 7]. The simulations were based on the molecular interaction model of Eckl et al. [17] that won the Fourth Industrial Fluid Properties Simulation Challenge in 2008 [18]. The model was optimized to experimental vapor pressure, saturated liquid density and heat of vaporization data as functions of temperature, which is a standard practice today. Its parameters and additional simulation details can be found in the appendix.

It should be pointed out that the accuracy of the EOS with respect to experimental data depends in the present context on two factors: (1) The quality of the molecular model that represents ethylene oxide here. (2) The representation quality of the simulation data by the EOS.

The investigation of factor (2) is the subject of this paper. Factor (1), the quality and efficient optimization of molecular models, is crucial when comparing molecular simulation data with experimental results, and this topic is perhaps the most discussed issue in the simulation community. It is generally accepted, however, that molecular models can, and often do, perform well for state points, properties and scenarios that were not considered during their optimization. This assumption is supported by our previous findings, in which numerous EOS correlations based on experimental data were compared with A^{r}_{mn} simulation results based on molecular models that were optimized exclusively to experimental VLE data [1].

6. Results

In order to validate the automatically generated EOS, its representation of the input data from molecular simulation was assessed. Furthermore, it was compared to experimental data from the literature, references can be found in Table 2, and to correlations thereof [19, 20]. This comparison was done on the basis of VLE properties, i.e. vapor pressure, saturated liquid density, saturated vapor density and heat of vaporization.

The agreement of the present EOS with the Helmholtz energy derivatives from molecular simulation can be seen in terms of the relative deviation plots in Figure 7. The symbols refer to the isotherms introduced in Figure 6. Figure 7 shows that for an increasing order of the derivatives the representation of the molecular simulation data gets worse, mostly due to the larger statistical uncertainties of the simulation data. A^{r}_{10} and A^{r}_{01} agree mainly within $\pm$ 2 %, whereas A^{r}_{20} and A^{r}_{02} agree mainly within $\pm$ 10 %. The other Helmholtz energy derivative (i.e. A^{r}_{11}) deviates within $\pm$ 4 %. Deviation Data for A^{r}_{30}, A^{r}_{21}, A^{r}_{12} are not shown, since their statistical uncertainties are rather high compared to the other derivatives. Which is why the EOS always meets A^{r}_{30}, A^{r}_{21}, A^{r}_{12} within their uncertainties. Note that the numerical values for the residual reduced Helmholtz energy A^{r}_{mn} are rather small, resulting in large relative deviations while the absolute deviations are small too, cf. Figure 8. It is apparent from the deviation plots that the generalized MBWR EOS performs quite well to fit the entire dataset.

Figure 9 shows the relative deviations in terms of the vapor pressure. The baseline in all VLE deviation plots represents the reference EOS by Thol et al. [4], which is more advanced and of higher quality than the one automatically generated in this work. It can be seen that the agreement between the present EOS and the experimental data below 260 K is poor. Between 290 and 470 K, the agreement gets more reasonable with deviations of under 2.5 % compared to the EOS by Thol et al. [4] Since the EOS created in this work is based exclusively on the molecular model by Eckl et al. [17], it coincides well with their simulation data. In Figure 10, comparisons for the saturated liquid density data are presented. In the range where most experimental data are available, i.e. from 220 to 320 K, the present EOS agrees within 1 % to the reference EOS [4] and the experimental data. At higher temperatures, the saturated liquid density is overestimated, yielding deviations of up to +5 % at 440 K.

Table 2: Experimental data for ethylene oxide from the literature. The units as given in the sources were converted into SI units and the temperature is given in terms of the international temperature scale of the 1990 standard (ITS-90). Data points calculated from an ancillary equation are marked with an asterisk.

Source	Year	Data points	Temperature range (K)	Pressure range (MPa)
Homogeneous density				
Lide [21] (1)	2005	1	273.15	0.101325
Walters & Smith [22] (1)	1952	81	294−428	0.006−3.448
Overall (3)		*82*	*273−428*	*0.006−3.448*
Vapor pressure				
Calado et al. [23]	1996	1	182.33	0.0001
Chao et al. [24]	1986	1	283.71	0.101325
Coles & Popper [25]	1950	17	273−305	0.067−0.221
Giauque & Gordon [26]	1949	14	223−286	0.004−0.108
Giles & Wilson [27]	2006	2	298−349	0.174−0.767
Gillespie et al. [28]	1985	2	283−299	0.101−0.174
Hess & Tilton [29]	1950	1	293.14	0.1462
Kistiakowsky & Rice [30]	1940	1	283.84	0.101325
Lide [21]	2005	3	283−284	0.101−0.100
Maass & Boomer [31]	1922	21	216−286	0.002−0.110
McDonald et al. [32]	1959	11	284−239	0.103−0.012
Mock & Smith [33]	1950	10	322−423	0.379−3.827
Olson [34]	1977	3	273−324	0.065−0.394
Frenkel et al. [19]*	2013	17	160−469	0.000−7.207
Walters & Smith [22]	1952	12	294−469	0.151−7.192
Overall (3)		*99*	*182−423*	*0.000−3.827*
Saturated liquid density				
Auwers [35]	1918	2	279.20	–
Comelli & Francesconi [36]	1991	11	288−304	–
Comelli & Francesconi [37]	1995	1	298.15	–
Comelli & Francesconi [38]	1996	2	298−314	–
Francesconi & Comelli [39]	1994	1	298.15	–
Francesconi & Comelli [40]	1995	1	298.15	–
Maass & Boomer [31]	1922	16	222−294	–
Olson [34]	1977	3	273−324	–
Perkinsen [41]	1893	1	280.15	–
Frenkel et al. [19]*	2013	17	160−469	–
Walters & Smith [22]	1952	12	294−469	–
Wurtz [42]	1863	1	273.15	–
Overall (3)		*51*	*222−469*	
Saturated vapor density				
Olson [34]	1977	3	273−324	–
Frenkel et al. [19]*	2013	20	377−469	–
Walters & Smith [22]	1952	12	294−469	–
Overall (3)		*15*	*273−469*	
Speed of sound				
Hurly [43]	2002	334	285−440	0.049−1.020
Overall (3)		*334*	*285−440*	*0.049−1.020*
Enthalpy of vaporization				
Cox & Pilcher [44]	1970	1	298.10	–
Rowley et al. [20] * (2)	2006	30	160−423	–
Giauque & Gordon [26]	1949	1	283.65	–
Lange & Dean [45]	1973	1	283.75	–
Lide [21]	2005	2	283−299	–
Matheson [46] *	1980	5	233−294	–
Reid et al. [47]	1977	1	283.66	–
Timmermans [48]	1965	1	288.10	–
Walters & Smith [22] *	1952	11	294−461	–
Washburn [49]	(1926−1933)	1	286.10	–
Yaws [50] *	1977	15	173−454	–
Overall (3)		*8*	*283−299*	

(1) One of the data points is located in the two-phase region and was thus neglected in the following.
(2) Calculated from the Clausius-Clapeyron equation.
(3) The overall values do not include any data derived from ancillary equations.

For the saturated vapor density, only very little experimental data are available, cf. Figure 11. From 270 to 420 K, the deviations of the present EOS compared to the equation of Thol et al. [4] are within ± 4 %. In the vicinity of the critical point the agreement is worse. Figure 12 shows the deviations with respect to the heat of vaporization. At temperatures below 335 K the experimental data and the EOS by Thol et al. [4] are underestimated by the present EOS, above 335 K they are overestimated. Furthermore, the ancillary equation from the DIPPR database is reproduced well at higher temperatures. The agreement between the simulation data by Eckl et al. [17] and the present EOS is consistently good. Deviation plots for density and speed of sound in the homogeneous fluid region can be found in Figures C.13 and C.14. At lower pressures, the present EOS and the one by Thol et al. [4] agree similarly well with the experimental data for the speed of sound and the homogeneous density. At higher pressures the representation of the present EOS gets worse, resulting in a maximum relative deviation of around -0.4 % for the speed of sound and around +3.0 % for the homogeneous density.

Conclusion

The cloud-based molecular simulation framework developed here allows for the automatic execution of large-scale simulation workflows across multiple, distributed computing resources. Simulations were concurrently deployed across two HPC systems, a Cray supercomputer and a large compute cluster. Carrying out a full workflow, i.e. defining the system, preparing and executing simulation runs and fitting the final EOS, is simplified to a few clicks in a user-friendly interface, thus enabling non-expert users to create fundamental EOS in a matter of days. The fully automatized generation of EOS on the basis of molecular modelling and simulation is a valuable process that is well supported by the powerful capabilities of cloud computing. This avoids most of the obstacles that currently confound the delivery, accessibility and usability of this non-service-oriented type of simulation on state of the art computing resources. According to the proposals of Maginn [8] the present work provides a contribution to making molecular simulation more accessible, hopefully encouraging more entities to use such techniques.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 609029 (FORTISSIMO). The simulations were carried out on the Cray XE6 (Hermit) at the High Performance Computing Center Stuttgart (HLRS) and on the OCuLUS cluster of the Paderborn Center for Parallel Computing (PC^2).

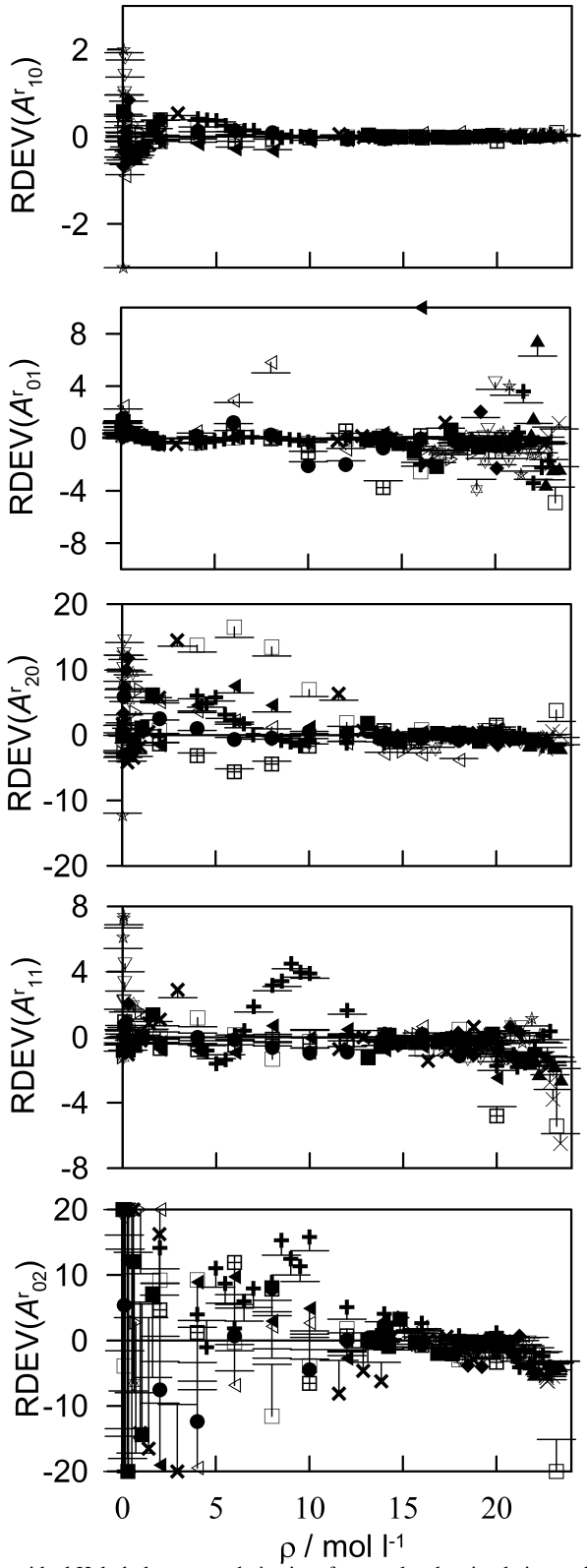

Figure 7: Relative deviation of the residual Helmholtz energy derivatives from molecular simulation to the present fundamental equation of state (baseline). Symbols refer to different isotherms.

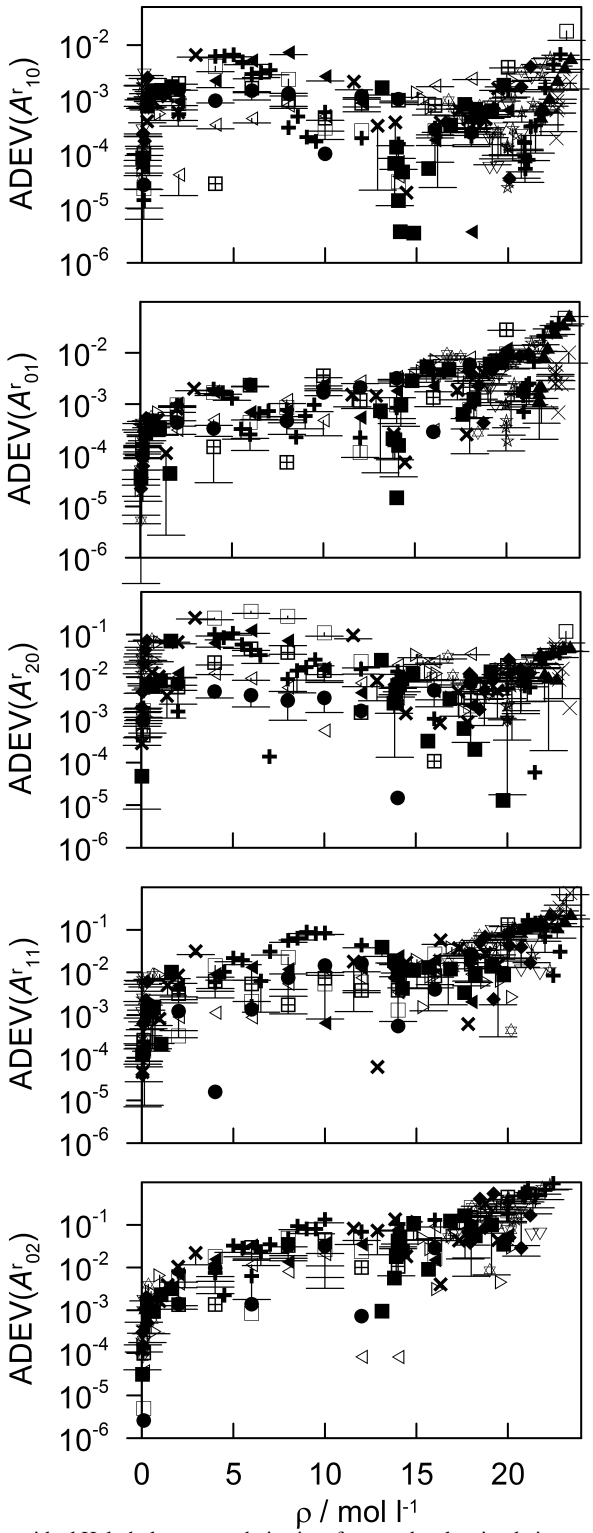

Figure 8: Absolute deviation of the residual Helmholtz energy derivatives from molecular simulation to the present fundamental equation of state. Symbols refer to different isotherms.

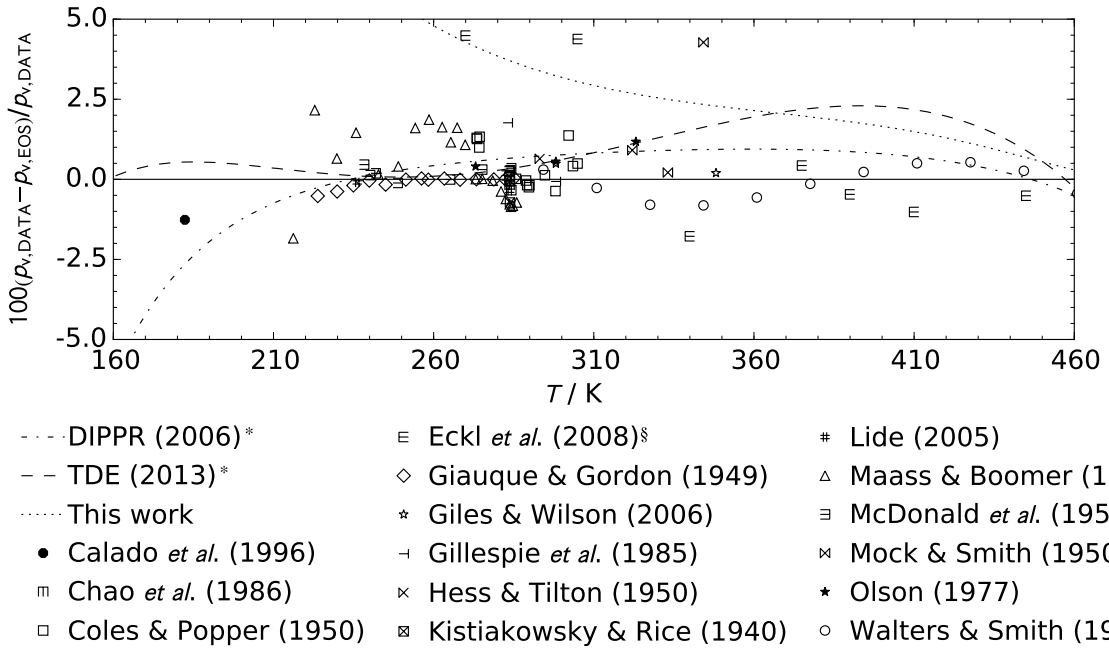

Figure 9: Comparison of the present fundamental equation of state with the one by Thol et al. [4] (baseline) and experimental data for the vapor pressure. The asterisk marks ancillary equations and the paragraph symbol indicates molecular simulation data by Eckl et al. [17]. The statistical uncertainty of the molecular simulation data increases gradually with decreasing temperature (0.6 % at 445 K, 1.6 % at 340 K and 17 % at 270 K).

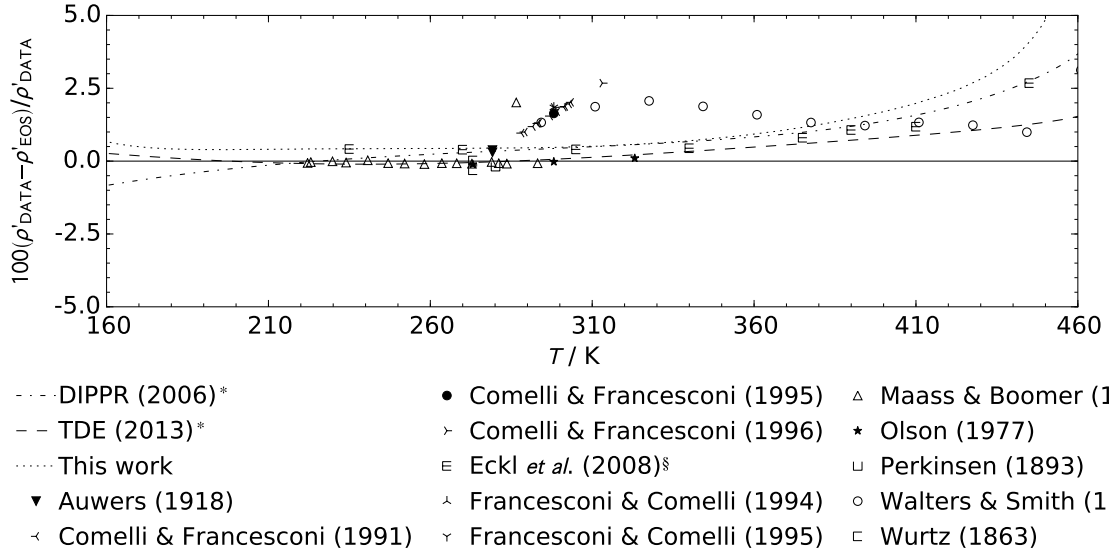

Figure 10: Comparison of the present fundamental equation of state with the one by Thol et al. [4] (baseline) and experimental data for the saturated liquid density. The asterisk marks ancillary equations and the paragraph symbol indicates molecular simulation data by Eckl et al. [17]. The statistical uncertainty of the molecular simulation data decreases gradually with decreasing temperature (0.5 % at 445 K, 0.04 % at 340 K and 0.02 % at 235 K).

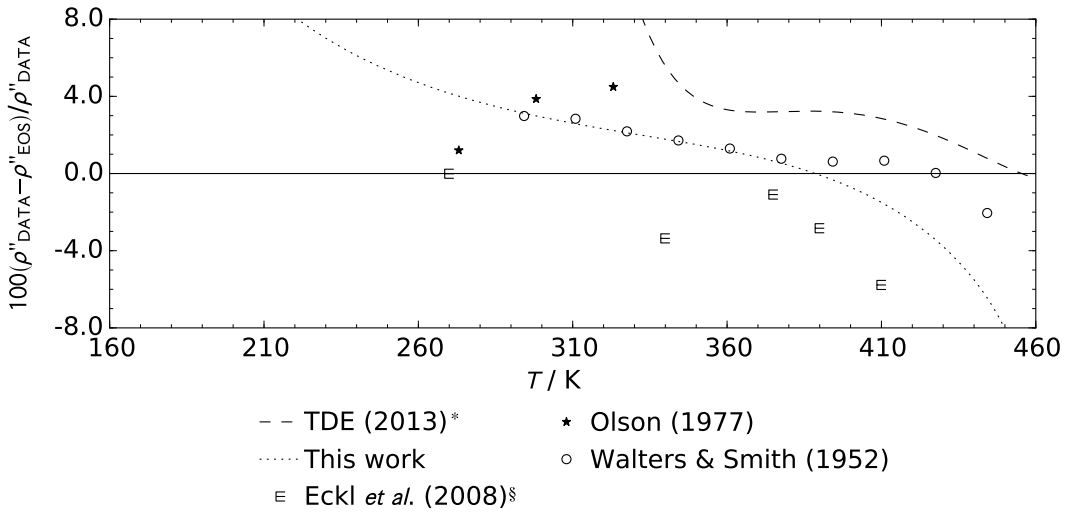

Figure 11: Comparison of the present fundamental equation of state with the one by Thol et al. [4] (baseline) and experimental data for the saturated vapor density. The asterisk marks an ancillary equation and the paragraph symbol indicates molecular simulation data by Eckl et al. [17]. The statistical uncertainty of the molecular simulation data decreases gradually with decreasing temperature (0.5 % at 410 K, 0.3 % at 340 K and 0.08 % at 270 K).

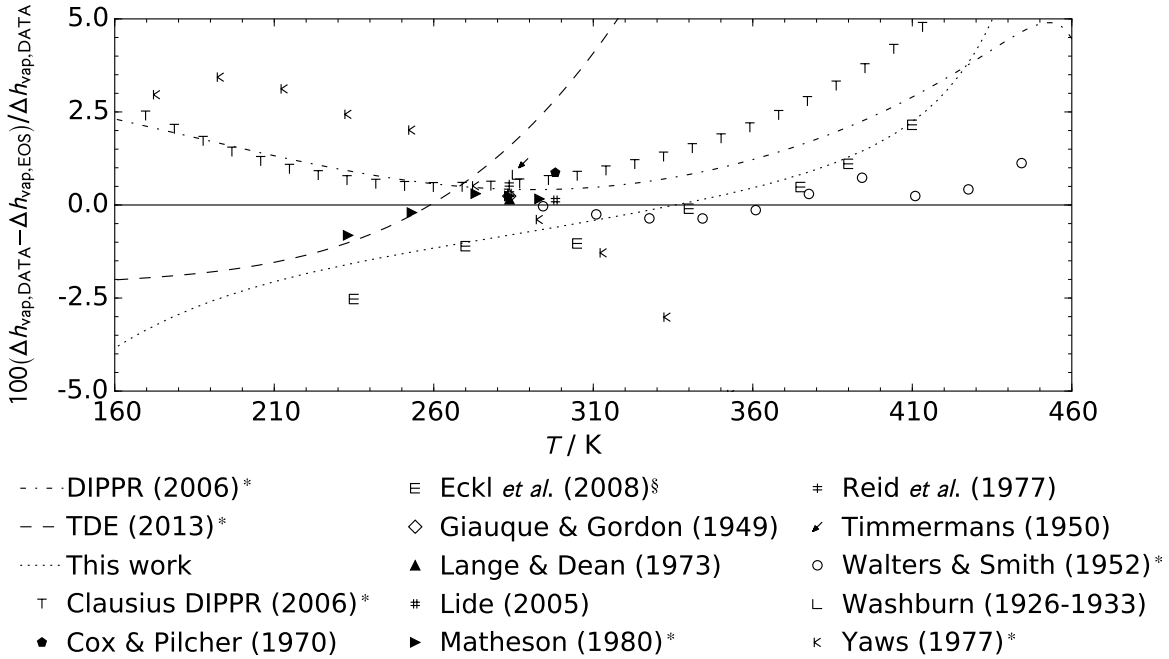

Figure 12: Comparison of the present fundamental equation of state with the one by Thol et al. [4] (baseline) and experimental data for the enthalpy of vaporization. The asterisk marks ancillary equations and the paragraph symbol indicates molecular simulation data by Eckl et al. [17]. The statistical uncertainty of the molecular simulation data decreases gradually with decreasing temperature (0.2 % at 410 K, 0.04 % at 340 K and 0.04 % at 235 K).

Appendix A. Simulation details

At each state point 864 particles were sufficiently equilibrated and then sampled for 2 million production cycles with NVT Monte Carlo simulations [16]. The cut-off radius was set to be half the length of the simulation box. Electrostatic long-range corrections were approximated by the reaction field method [51].

Appendix B. Molecular model

The coordinates and parameters of the molecular model for ethylene oxide are given in Table B.3. The model consists of three Lennard-Jones sites and one point dipole and was optimized to experimental data for vapor pressure, saturated liquid density and enthalpy of vaporization.

Table B.3: Sigma (σ) and epsilon (ϵ) are the length and energy parameters of the Lennard-Jones (LJ) potential, respectively. k_B is the Boltzmann constant. μ denotes the dipole moment of the point dipole. All coordinates are given in a principal axes system with respect to the center of mass. The orientation of the point dipole is defined with Euler angles: φ is the azimuthal angle with respect to the x-y-plane and θ is the inclination angle with respect to the z-axis.

Site	x/Å	y/Å	z/Å	σ/Å	ϵ/k_B/K	θ/deg	φ/deg	μ/D
CH_2	0.78	0	-0.48431	3.5266	84.739			
CH_2	-0.78	0	-0.48431	3.5266	84.739			
O	0	0	0.73569	3.0929	62.126			
Dipole	0	0	0			0	0	2.459

Appendix C. Additional data

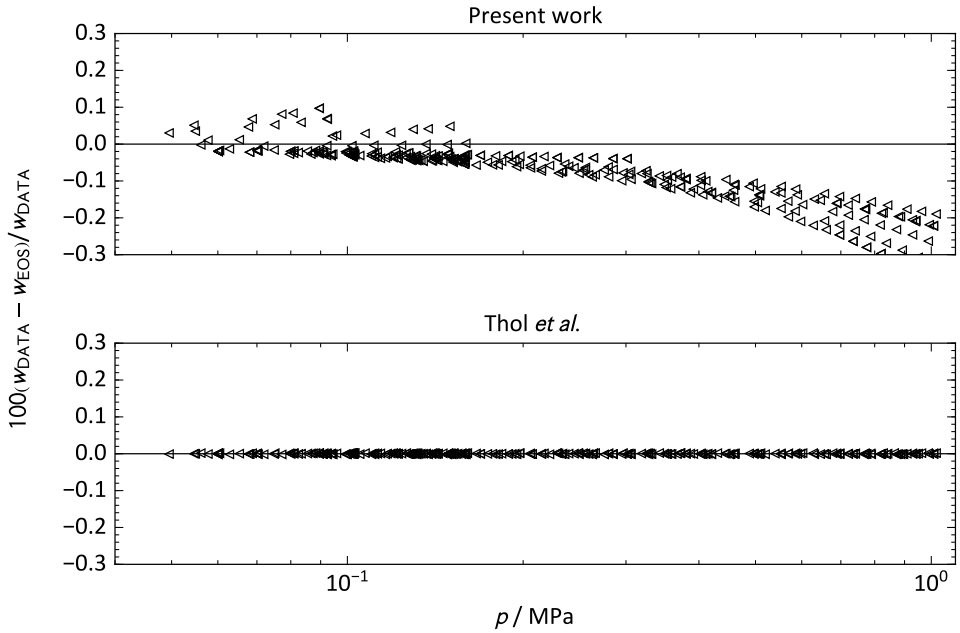

Figure C.13: Comparison of the present fundamental equation of state (top) with the one by Thol et al. [4] (bottom) for the speed of sound w. Experimental speed of sound data by Hurly [43] ($\triangleleft$) are plotted.

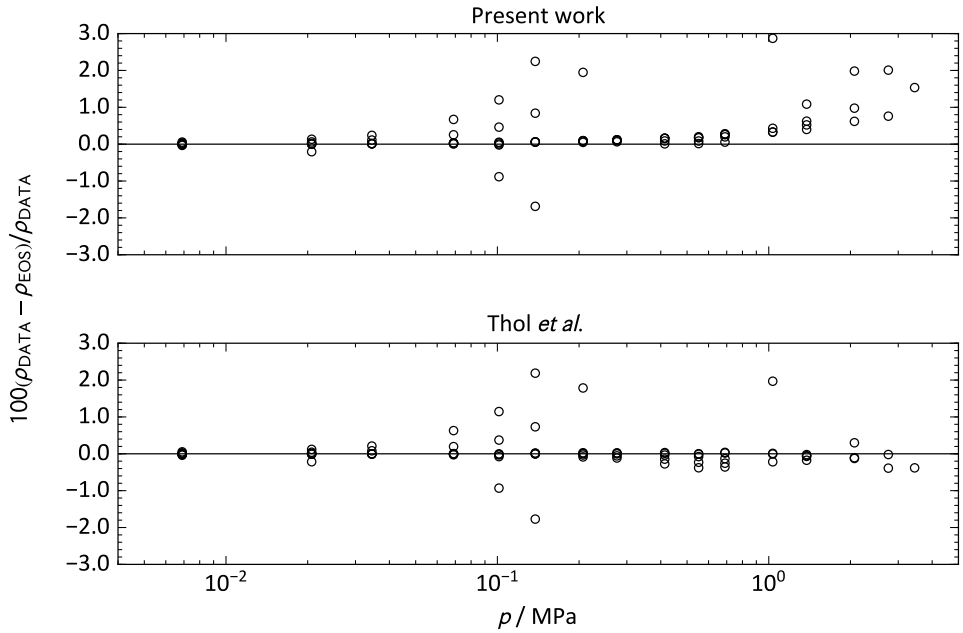

Figure C.14: Comparison of the present fundamental equation of state (top) with the one by Thol et al. [4] (bottom) for density ρ in homogeneous fluid states. Experimental data by Walters and Smith [22] (○) are plotted.

[1] G. Rutkai, M. Thol, R. Lustig, R. Span, J. Vrabec, The Journal of Chemical Physics 139 (2013) 041102.

[2] G. Rutkai, J. Vrabec, Journal of Chemical & Engineering Data 60 (2015) 2895–2905.

[3] M. Thol, G. Rutkai, R. Span, J. Vrabec, R. U. Lustig, International Journal of Thermophysics 36 (2015) 25–43.

[4] M. Thol, G. Rutkai, A. Köster, M. Kortmann, R. Span, J. Vrabec, Chemical Engineering Science 121 (2015) 87–99, and 134 (2015) 887–890.

[5] R. Span, Multiparameter Equations of State: An Accurate Source of Thermodynamic Property Data, Springer Verlag, Berlin, 2000.

[6] R. Lustig, Molecular Simulation 37 (2011) 457–465.

[7] R. Lustig, Molecular Physics 110 (2012) 3041–3052.

[8] E. J. Maginn, AIChE Journal 55 (2009) 1304–1310.

[9] 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, Computer Physics Communications 185 (2014) 3302–3306.

[10] P. J. Mohr, B. N. Taylor, D. B. Newell, Reviews of Modern Physics 84 (2012) 1527–1605.

[11] E. W. Lemmon, R. T. Jacobsen, Journal of Physical and Chemical Reference Data 34 (2005) 69–108.

[12] R. Span, W. Wagner, International journal of thermophysics 24 (2003) 1–39.

[13] L. Sun, J. Ely, International Journal of Thermophysics 26 (2005) 705–728.

[14] R. T. Jacobsen, R. B. Stewart, Journal of Physical and Chemical Reference Data 2 (1973) 757–922.

[15] J. Hust, R. McCarty, Curve-fitting techniques and applications to thermodynamics, Cryogenics 7 (1967) 200–206.

[16] D. Frenkel, B. Smit, Understanding Molecular Simulation: From Algorithms to Applications, Academic Press, Elsevier, San Diego, 2002.

[17] B. Eckl, J. Vrabec, H. Hasse, Fluid Phase Equilibria 274 (2008) 16–26.

[18] 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 274 (2008) 2–9.

[19] M. Frenkel, R. D. Chirico, V. Diky, K. Kroenlein, C. D. Muzny, A. F. Kazakov, J. W. Magge, I. M. Abdulagatov, E. W. Lemmon, NIST Standard Reference Database 103b: NIST Thermo–Data Engine – Pure Compounds, Binary Mixtures, Reactions, Version 8.0, National Institute of Standards and Technology, Standard Reference Data Program, Gaithersburg, 2013.

[20] R. L. Rowley, W. V. Wilding, J. L. Oscarson, Y. Yang, , N. A. Zundel, T. E. Daubert, R. P. Danner, DIPPR Data Compilation of Pure Compound Properties: Design Institute for Physical Properties, AIChE, New York, 2006.

[21] D. R. Lide, CRC Handbook of Chemistry and Physics, 86 ed., Taylor and Francis, London, 2005.

[22] C. J. Walters, J. M. Smith, Chemical Engineering Progress 48 (1952) 337–343.

[23] J. C. G. Calado, U. K. Deiters, E. J. M. Filipe, The Journal of Chemical Thermodynamics 28 (1996) 201–207.

[24] J. Chao, K. R. Hall, K. N. Marsh, R. C. Wilhoit, Journal of Physical and Chemical Reference Data 15 (1986) 1369–1436.

[25] K. F. Coles, F. Popper, Industrial and Engineering Chemistry 42 (1950) 1434–1438.

[26] W. F. Giauque, J. Gordon, Journal of the American Chemical Society 71 (1949) 2176–2181.

[27] N. F. Giles, G. M. Wilson, Journal of Chemical & Engineering Data 51 (2006) 1954–1962.

[28] P. C. Gillespie, J. R. Cunningham, G. M. Wilson, AIChE Symposium Series 81 (1985) 26–40.

[29] L. G. Hess, V. V. Tilton, Industrial and Engineering Chemistry 42 (1950) 1251–1258.

[30] G. B. Kistiakowsky, W. W. Rice, The Journal of Chemical Physics 8 (1940) 618–622.

[31] O. Maass, E. H. Boomer, Journal of the American Chemical Society 44 (1922) 1709–1728.

[32] R. A. McDonald, S. A. Shrader, D. R. Stull, Journal of Chemical & Engineering Data 4 (1959) 311–313.

[33] J. E. Mock, J. M. Smith, Industrial and Engineering Chemistry 42 (1950) 2125–2128.

[34] J. D. Olson, Journal of Chemical & Engineering Data 22 (1977) 326–329.

[35] K. V. Auwers, Justus Liebig's Annalen der Chemie 415 (1918) 98–168.

[36] F. Comelli, R. Francesconi, Journal of Chemical & Engineering Data 36 (1991) 382–383.

[37] F. Comelli, R. Francesconi, Journal of Chemical & Engineering Data 40 (1995) 28–30.

[38] F. Comelli, R. Francesconi, Journal of Chemical & Engineering Data 41 (1996) 101–104.

[39] R. Francesconi, F. Comelli, Journal of Chemical & Engineering Data 39 (1994) 106–107.

[40] R. Francesconi, F. Comelli, Journal of Chemical & Engineering Data 40 (1995) 512–514.

[41] W. H. Perkin, Journal of the Chemical Society, Transactions 63 (1893) 488–491.

[42] A. Wurtz, Annales de Chimie et de Physique 55 (1859) 400–432.

[43] J. J. Hurly, International Journal of Thermophysics 23 (2002) 667–696.

[44] J. D. Cox, G. Pilcher, Thermochemistry of Organic and Organometallic Compounds, Academic Press, New York, 1970.

[45] N. A. Lange, J. A. Dean, Lange's Handbook of Chemistry, eleventh ed., McGraw-Hill, New York, 1973.

[46] Matheson, Matheson Gas Data Book, sixth ed., Lyndhurst, New Jersey, 1980.

[47] R. C. Reid, J. M. Prausnitz, T. K. Sherwood, The Properties of Gases and Liquids, third ed., McGraw-Hill, New York, 1977.

[48] J. Timmermans, Physico-Chemical Constants of Pure Organic Substances: Vol. II, Elsevier, New York, 1965.

[49] E. W. Washburn, International Critical Tables of Numerical Data, Physics, Chemistry, and Technology: 7 Volumes + Index, McGraw-Hill, New York, 1926–1933.

[50] C. L. Yaws, Physical Properties: A Guide to the Physical, Thermodynamic, and Transport Property Data of Industrially Important Chemical Compounds, McGraw-Hill, New York, 1977.

[51] J. A. Barker, R. O. Watts, Molecular Physics 26 (1973) 789–792.