Check for
updates

Planning of Measurement Series for Thermodynamic Properties Based on Optimal Experimental Design

Ophelia Frotscher[1] · Roland Herzog[2] · Markus Richter[1] 📷

Abstract

Decreasing the time required for accurate thermodynamic property measurements is extremely desirable for model development that can respond to the needs of science and industry within a short time frame. Here, we demonstrate the application of optimal experimental design to measurements of thermodynamic properties. The technique is explored using the fitting of a Schilling-type equation from published (p, ρ, T)-measurements of ethylene glycol and propylene glycol. The analysis shows that a fixed-exponent fit using (p, T)-measurements along the five most informative isotherms produces models of relative density errors comparable to those obtained using the data along all investigated isotherms, i.e., eight or nine. It is also argued that a calculation of optimal isotherms prior to the measurement series can further increase the precision at no additional experimental effort.

Keywords Correlation · Density · Ethylene glycol · Measurement · Optimal experimental design

1 Introduction

For scientific, but particularly for industrial applications, it is absolutely valuable to provide equations of state (EOS) that accurately describe the thermodynamic behavior of the concerned fluid substances within a narrow time frame. However, accurate

✉ Markus Richter
m.richter@mb.tu-chemnitz.de

Ophelia Frotscher
ophelia.frotscher@mb.tu-chemnitz.de

Roland Herzog
roland.herzog@mathematik.tu-chemnitz.de

[1] Technische Universität Chemnitz, Faculty of Mechanical Engineering, Professorship Applied Thermodynamics, 09107 Chemnitz, Germany

[2] Technische Universität Chemnitz, Faculty of Mathematics, Professorship Numerical Mathematics (Partial Differential Equations), 09107 Chemnitz, Germany

thermodynamic models are usually based on reliable experimental data, thus, the development can require quite a long time since accurate measurements over wide temperature and pressure ranges are typically a time-consuming endeavor. Here, we demonstrate the advantages of applying optimal experimental design (OED) in the process of planning measurement series for thermodynamic properties with the goal to deliberately reduce the experimental effort, thus, decreasing the overall time for model development. Therefore, the issue with typical experimental design and the approach of OED will be explained using our density data of ethylene glycol published by Yang *et al.* [1] and the density data of propylene glycol measured with the same instrument published by Sampson *et al.* [2].

The (p, ρ, T) behavior of ethylene glycol was investigated over the temperature range from $T = (283.3$ to $393.1)$ K at pressures from $p = (4.8$ to $100.1)$ MPa utilizing a high-pressure vibrating-tube densimeter; as shown in Fig. 1 , a total of 89 (p, ρ, T) data points were studied with a combined expanded uncertainty $(k = 2)$ of 1.57 kg·m^{-3} (equivalent to a max. relative uncertainty of 0.151 %). To model the experimental data, Yang *et al.* [1] fitted two empirical Schilling-type correlation equations [3]: one of the same form (same number of terms, same exponents) as for propylene glycol in the work of Sampson *et al.* [2], and one using the "artificial intelligence powered" software tool Eureqa [4, 5]. Eureqa is one of many tools that perform machine learning-based symbolic regression. It simultaneously finds the functional form of a model and identifies the model parameters. The first approach was motivated by the assumed thermodynamic similarity of both substances, and the second approach served to investigate the applicability of symbolic regression and machine learning for the optimization of correlation equations. Section 2 of the present paper provides a small overview of different
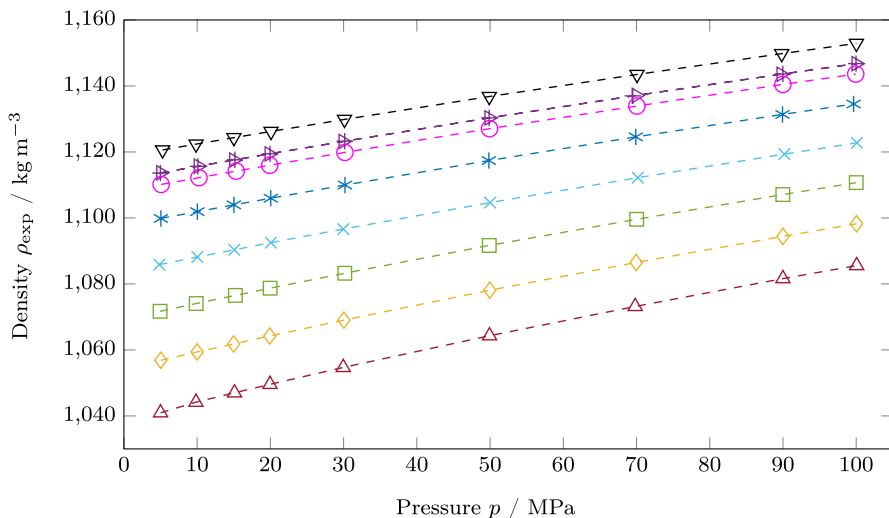

Fig. 1 Experimental densities (symbols) and interpolated densities (dashed lines) measured by Yang *et al.* [1]. $\triangledown$, $T \approx 283.32$ K; $+$, $T \approx 293.14$K; $\triangleright$, $T \approx 293.14$ K; $\bigcirc$, $T \approx 298.18$ K; $*$, $T \approx 313.13$ K; $\times$, $T \approx 333.09$ K; $\square$, $T \approx 353.13$ K; $\diamond$, $T \approx 373.22$ K; $\triangle$, $T \approx 393.09$ K

types of correlation models for liquid-phase densities to support the understanding of their characteristics.

In the present study, we illustrate how OED could have been used for the planning of the measurement series conducted by Yang *et al.* [1] in order to reduce the amount of density measurements of ethylene glycol needed to reliably adjust the parameters of a Schilling-type equation. To confirm our results regarding the benefits of OED, the calculations were also applied to measured densities of propylene glycol (combined expanded uncertainty ($k = 2$): 1.78 kg·m^{-3}) [2]. In spite of its capabilities to select measurements with the highest information content, OED is seldom applied in the field of thermodynamic property research. We conjecture that the primary reasons for this are:

– a lack of knowledge about OED technologies,
– the lack of software tools for easy application of OED,
– the requirement to specify the underlying model beforehand.

For a detailed overview of the use of OED in different areas of chemical and thermal engineering research, we refer the reader to Francheschini and Macchietto [6]. With respect to thermodynamic property research, we point out the work of Bardow and colleagues [7, 8], which deals with the experimental characterization of liquid–liquid equilibria utilizing OED. In contrast to those investigations, the present work focuses on the temperature and pressure values for density measurements, with the aim to minimize the parameter variance of the correlation model. The method described here can be used equivalently for measurements of other thermodynamic and transport properties (e.g., speed of sound, specific heat capacity, viscosity, etc.).

When changing the isotherm in a usual measurement series, establishing thermal equilibrium with a typical (p, ρ, T) apparatus is rather time-consuming compared to setting new pressures along an isotherm. For this reason, we decided to use OED to select a subset of the most informative isotherms. OED can be imagined as an in situ tool for linear or nonlinear modeling. In the latter, case, the following sequential workflow is employed:

(a) choose an initial set of (p, T) state points to be studied (e.g., two isotherms with five pressure values each),
(b) conduct the measurements (e.g., density),
(c) fit the model (e.g., the Schilling-type model),
(d) answer the following questions:

　– Can the model reproduce the measured data with sufficient accuracy (e.g., within the experimental uncertainty)?
　– Can a significant amount of information be gained from additional measurements?

(e) If the latter is not the case, the experiment is terminated. Otherwise, select the next most relevant isotherm using OED (3) and proceed with 1.

In this paper, we utilize in a first step, linear OED to select a subset of isothermal experiments from two given sets of measurements, which are already available. We use the measurements selected to create new models for ethylene glycol and propylene glycol based on the same functional forms as used before. The results are compared to the existing models (4).

2 Modeling Approach

For simulations in process engineering, accurate and easily applicable models that describe the real fluid behavior are needed. Considering the lack of knowledge about more complex compounds at the molecular level, empirical equations of state have become an established way of modeling thermodynamic behavior. In industrial applications, the empirical Tait equation (1888) is still widely used to model densities of liquids [9]. However, a contemporary approach to model liquid-phase densities was developed by Schilling *et al.* [3] (Table 1). This is an equation in form of polynomial terms; this type of terms is also used within fundamental equations of state.

For the modeling of fundamental equations of state, which provide the possibility to derive all thermodynamic properties by differentiation of these equations, functional forms such as polynomial and exponential terms are involved [10, 11]. In 1989, within the scope of developing fundamental equations, Setzmann and Wagner [10, 12] established the modeling tool OPTIM, which combines a modified step-wise regression analysis based on a "*bank of terms*" with elements of evolutionary optimization methods. Nowadays, this tool is mostly unused. However, it was applied to model the density of liquid *n*-heptane, *n*-nonane, 2,4-dichlorotoluene and bromobenzene in the temperature range from 233.15 to 473.15 K at pressures up to 30 MPa by Schilling *et al.* [3]. Due to its good representation of liquid-phase experimental data, further works, e.g., of Sommer *et al.* [13], Sampson *et al.* [2] and Yang *et al.* [1], also relate to this equation. Table 4 (M1 and M1.1) contains the characteristics and Table 5 (M1 and M1.1) provides the parameter values of the published models for ethylene glycol and propylene glycol, which are used in this work. The so-called Schilling-type equation has the following functional form:

$$\frac{\rho}{\rho_0} = \sum_{j=1}^{I_{\text{Pol}}} n_j \, \sigma^{t_j} \pi^{p_j} \tag{2.1}$$

$$\sigma = (T/T_0 - 1) \text{ and } \pi = (p/p_0 + 1) \tag{2.2}$$

The parameters σ and π represent the reduced temperature and the reduced pressure, respectively, while T_0, p_0 and ρ_0 are chosen by the measured values and I_{Pol} defines the number of terms. In Sampson *et al.* [2], the parameters were set to $T_0 = 150$ K, $p_0 = 100$ MPa and $\rho_0 = 1000$ kg·m^{-3}, and n_j, t_j and p_j were fitted to the experimental data of propylene glycol setting the number of terms I_{Pol} 8. With this model, the measured densities for propylene glycol can be reproduced within a maximum error of $\varepsilon_R = 0.015$ %. Due to the similarity between both substances, the same setup was

used to fit the parameters n_j for ethylene glycol in Yang *et al*. [1], keeping the exponents t_j and p_j fixed (see Tables 4 (M1) and 5). The experimental data of ethylene glycol is represented within a maximum error of $\varepsilon_R = 0.025\,\%$.

Most current modeling approaches, especially for nonlinear models, have a high development effort and the requirement of thorough background knowledge in common. In the work of Yang *et al*., also a new approach utilizing Eureqa (Table 1) was employed to investigate the suitability of machine learning-based symbolic regression for nonlinear EOS modeling [4, 5]. The form of terms was chosen as polynomials to create a Schilling-type equation 2.1 and 2.2 equivalent to the existing model, while the aim was to reduce the number of terms, to fit the exponents and to stay within an adequate uncertainty (for model characteristics see Table 4 (M1.1) and for model parameters see Table 5 (M1.1)). This model is included in the evaluation, on the one hand to better interpret the deviations between modeled and measured densities and on the other hand to better illustrate the influence of the functional form on the extrapolation behavior. In addition to the less complex functional form, a lower maximum error of $\varepsilon_R = 0.020\,\%$ and a better extrapolation behavior are achieved.

3 Background on Optimal Experimental Design

Optimal experimental design is a technique to select experiments which are most informative about the unknown parameters of a given model. We refer the interested reader to Pázman, Uciński and Atkinson *et al*. [14–16] for a comprehensive background. We wish to emphasize that OED differs from the statistical design of experiments, where one seeks, e.g., space-filling experiments by latin hypercube sampling. Here, we use OED to reduce the experimental effort required to identify the parameters n_j, $j = 1, \ldots, I_{Pol}$, in Eqs. 2.1 and 2.2 with exponents t_j and p_j fixed. In this case, the dependent variable ρ/ρ_0 is a linear function of the parameters, which simplifies the subsequent description. To be specific, we deduce from Eqs. 2.1 and 2.2 that each triple of measured data (p_i, ρ_i, T_i), after conversion to reduced

Table 1 Features of different EOS modeling approaches for liquid-phase densities (Tait equation [9], Schilling-type equation [3], Eureqa modeling [4, 5])

| EOS type	Features
Tait	Well-established model, widely used for liquid densities
	Predefined model form
	Requires custom fitting algorithm
Schilling	User-specified number of terms
	Reduces to linear fitting when exponents are fixed
	Requires custom fitting algorithm when exponents are unknown
Eureqa	Intuitive software handling
	Symbolic regression based on genetic algorithms
	Comparing complexity and uncertainty
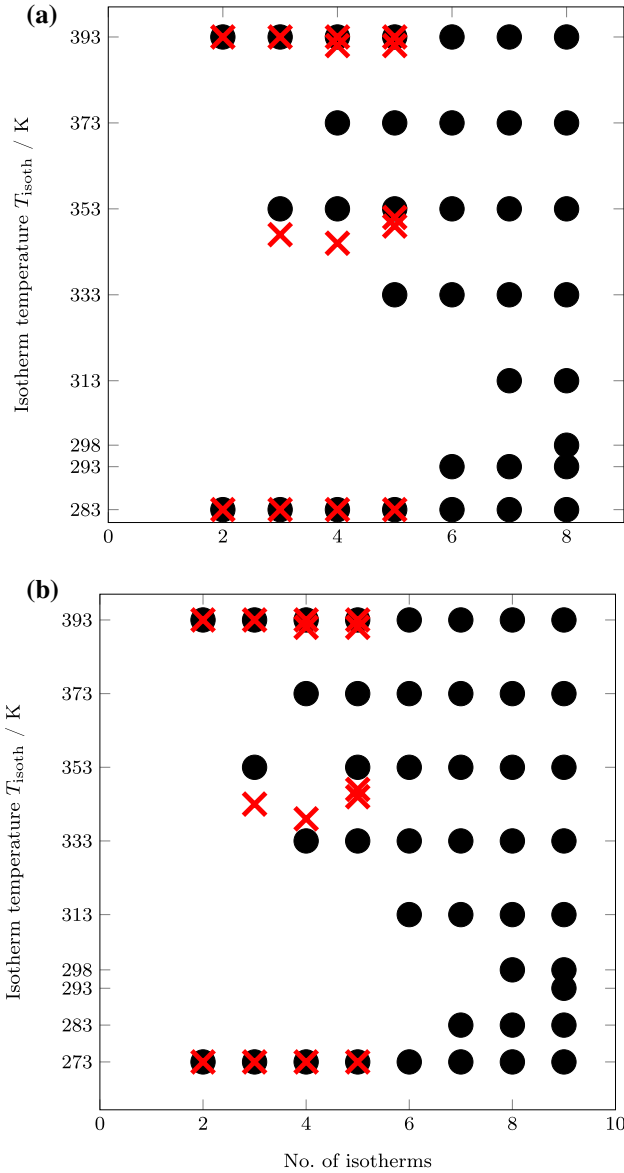
Fig. 2 Selection of the best isotherms • from the measured isotherms and × from a free choice a) for ethylene glycol between $T = (283 \text{ to } 393)$ K and b) for propylene glycol between $T = (273 \text{ to } 393)$ K, within a raster of 2 K using the A-criterion (Eq. 3.2) to minimize the parameter uncertainty

quantities $(\pi_i, \rho_i/\rho_0, \sigma_i)$, contributes a prediction of the form $\rho_i \approx \rho_0 j(\pi_i, \sigma_i) n$, where $n = (n_1, \dots, n_8)^{\mathrm{T}} \in \mathbb{R}^8$ denotes the parameter vector and

$$j(\pi_i, \sigma_i) = \begin{bmatrix} \sigma_i^{t_1} \pi_i^{p_1} & \cdots & \sigma_i^{t_8} \pi_i^{p_8} \end{bmatrix} \tag{3.1}$$

is the so-called elementary Jacobian associated with the ith measurement. Following the theory of OED, the information content of a single measurement is expressed through the elementary Fisher information matrix (FIM) $I_i = j(\pi_i, \sigma_i)^{\mathrm{T}} j(\pi_i, \sigma_i) \in \mathbb{R}^{8 \times 8}$. Each elementary FIM is a symmetric, positive semi-definite rank–1 matrix. Moreover, since we can assume individual measurements to be statistically independent, the FIM associated with a series of experiments is obtained as $I = \sum_i I_i$.

It is common to apply a scalar objective function, which converts the FIM associated with any collection of experiments into a single number and, thus, allows a comparison with any other collection of experiments. We utilize here the A-criterion

$$\Psi_A(I) = \operatorname{trace}(I^{-1}) = \sum_{j=1}^{8} \frac{1}{\lambda_j}, \tag{3.2}$$

where λ_j is the jth eigenvalue of I. Clearly, at least eight individual measurements are required to render the FIM positive definite and the criterion $\Psi_A(I)$ finite. In this case, the value of $\Psi_A(I)$ is proportional to the sum of the squared semi-axes of confidence ellipsoids in the 8-dimensional parameter space. Consequently, we seek to minimize $\Psi_A(I)$, possibly subject to constraints on the experimental budget, in order to maximize the information content of the collection of experiments selected and simultaneously minimize parameter variation in the face of measurement errors.

As we argued in the introduction, it is advantageous from a practical point of view to take measurements along an isotherm. In the following section, we will therefore optimize over experiments each of which comprises several measurements obtained by varying the pressure along an isotherm.

4 Numerical Results

The approach described in Sect. 3 is used to optimize the measurement series by maximizing the information contained in a set of eight isothermal measurements at approximate temperatures $T = (283, 293, 298, 313, 333, 353, 373, 393)$ K drawn from the experiments conducted by Yang et al. [1]. The measurement plan for each isotherm consisted of nine pressure values, approximately $p = (5, 10, 15, 20, 30, 50, 70, 90, 100)$ MPa. Table 2 and Fig. 2a show the most informative selection of isotherms (black dots) to fit the parameters $n_1, \ldots, n_8$ using a varying number of isotherms. As expected, the value of the objective (Eq. 3.2) decreases as we allow more isotherms to be included. However, when transitioning from five to six isotherms, the further decrease in the objective is small compared to the previous steps. For propylene glycol, measured at approximate temperatures $T = (273, 283, 293, 298, 313, 333, 353, 373, 393)$ K, Fig. 2b and Table 3 show a similar selection of best isotherms and for the decay of the objective function values. In this case, the

measurement plan for each isotherm consisted of eight pressure values, approximately at $p = (5, 10, 15, 20, 30.5, 50.5, 71, 91)$ MPa.

We aim to compare the accuracies of the existing models, see Table 4 (M1), with new models of the same functional form, where the parameter $n_1, \ldots, n_8$ of Eqs. 2.1 and 2.2 are fitted using only the five best isotherms, see Table 4 (M2). To this end, we show in Fig. 3a and c the relative deviations of calculated densities from experimental values as a function of pressure along all eight isotherms. It can be seen that using only five out of eight isotherms for the fitting process yields a greatest deviation only larger by approximately 0.003 %. The same approach is used to fit a new model for propylene glycol to the best five isotherms, see the lower part of Table 5 (M2). With an absolute difference in the greatest deviation of 0.0015 % for propylene glycol between the models M1 and M2 (Fig. 4a and b), the result for OED is even more promising than for ethylene glycol.

Next, we investigate the information gain obtained by allowing the temperatures of the isothermal experiments to be chosen freely within the interval from $T = 283$ K to $T = 393$ K for ethylene glycol and $T = 273$ K to $T = 393$ K for propylene glycol. To this end, we introduce an equidistant grid of possible temperature values with a spacing of 2 K; the pressure values for each group of isothermal experiments are the same as measured. The optimal selections of up to five isotherms can be found in Fig. 2a and the bottom half of Table 2 for ethylene glycol as well as in Fig. 2b and the bottom half of Table 3 for propylene glycol.

By comparing both selection approaches, the similarity of temperatures for the best two and three isotherms as well as the difference between those for the four and five best isotherms are noticeable. This leads us to expect an improvement of the model with a fit to the free best five isotherms. However, due to the missing measured values, no comparable model can be developed for this case.

Table 2 Selection of the best isotherms for different numbers of isotherms and the objective value (when fitted to that number of isotherms) for ethylene glycol

No. of isotherms	Selection from the pool of measured isotherms / K	Objective value
2	393.09, 283.32	23782.87
3	393.09, 283.32, 353.13	120.32
4	393.09, 283.32, 353.13, 373.22	94.40
5	393.09, 283.32, 353.13, 373.22, 333.09	78.82
6	393.09, 283.32, 353.13, 373.22, 333.09, 293.14	73.07
7	393.09, 283.32, 353.13, 373.22, 333.09, 293.14, 313.13	68.36
8	393.09, 283.32, 353.13, 373.22, 333.09, 293.14, 313.13, 298.18	65.77
No. of isotherms	Free selection of isotherms within the investigated temperature range / K	Objective value
2	283.00, 393.00	23113.13
3	283.00, 347.00, 393.00	119.83
4	283.00, 345.00, 391.00, 393.00	85.98
5	283.00, 349.00, 351.00, 391.00, 393.00	69.22

Table 3 Selection of the best isotherms for different numbers of isotherms and the objective value (when fitted to that number of isotherms) for propylene glycol

No. of isotherms	Selection from the pool of measured isotherms / K	Objective value
2	272.73, 392.95	37026.60
3	272.73, 392.95, 352.99	145.84
4	272.73, 392.95, 373.34, 333.05	107.35
5	272.73, 392.95, 352.99, 373.34, 333.05	90.86
6	272.73, 392.95, 352.99, 373.34, 333.05, 313.12	83.26
7	272.73, 392.95, 352.99, 373.34, 333.05, 313.12, 283.18	78.56
8	272.73, 392.95, 352.99, 373.34, 333.05, 313.12, 283.18, 298.12	75.55
9	272.73, 392.95, 352.99, 373.34, 333.05, 313.12, 283.18, 298.12, 293.18	73.47
No. of isotherms	Free selection of isotherms within the investigated temperature range / K	Objective value
2	273.00, 393.00	35604.70
3	273.00, 343.00, 393.00	144.11
4	273.00, 339.00, 391.00, 393.00	98.78
5	273.00, 345.00, 347.00, 391.00, 393.00	80.83

Instead, we investigate the importance of the functional form by fitting a new model to the best five measured isotherms, where also the exponents are optimized, see Fig. 3d, Tables 4(M3) and 5(M3). Here, even with five isotherms, the maximum deviations can be reduced significantly. However, for propylene glycol, the model M3 shows large deviations for the isotherms not selected (greatest deviation 0.0431 %) compared to the deviation of the selected isotherms (greatest deviation 0.0046 %) (Fig. 4c) and compared to the model published by Sampson *et al*. [2] (Tables 4 and 5, propylene glycol - M1). This behavior is in conjunction with the conclusion of Yang *et al*. [1] concerning the Eureqa model (Fig. 3b), which demonstrates the importance of a comprehensive model optimization including the functional form. When fitting the exponents, the model is not linear in the parameters anymore, which affects the selection of the best isotherms. To further exploit the potential of free exponent modeling, we plan to employ the sequential OED approach, described at the end of Sect. 1, for calculating the necessary isotherms based on nonlinear models in future work.

For a more comprehensive model examination between the modeling approaches, the extrapolation behavior was investigated taking the example of ethylene glycol; this was first done with an optimization of the functional form utilizing Eureqa, as shown in Table 4 with model M1.1, and second with parameter identification for a given functional form, as shown in Table 4 with models M1, M2 and M3. Therefore, the models were compared to an unpublished fundamental equation of state in form of the Helmholtz energy as implemented in REFPROP version 10.0 [17]. This model is valid within the following limits:

Table 4 Characteristics of the different Schilling type equations 2.1 and 2.2 (models M1, M1.1, M2 and M3) for ethylene glycol and propylene glycol used in this work

	Source	Development	Data points
M1	Ethylene glycol: Yang *et al*. [1]; propylene glycol: Sampson *et al*. [2]	8 terms; ethylene glycol: fitting of coefficients; exponents same as for propylene glycol; propylene glycol: fitting of coefficients and exponents (integers)	All data points, incl. reference measurements; ethylene glycol: 8 isotherms 9 pressures each; propylene glycol: 9 isotherms, 8 pressures each
M1.1	Yang *et al*. [1] (only for ethylene glycol)	Symbolic regression with Eureqa (optimized number of terms (7); parameter identification for linear coefficients and exponents (integers))	All data points, incl. reference measurements; 8 isotherms, 9 pressures each
M2	Developed within the scope of this work	Number of terms and exponents same as M1, fitting of coefficients	5 most informative isotherms; ethylene glycol: 9 pressures each; propylene glycol: 8 pressures each
M3	Developed within the scope of this work	Number of terms same as for M1, fitting of coefficients and exponents (real numbers)	5 most informative isotherms; ethylene glycol: 9 pressures each; propylene glycol: 8 pressures each

– temperature: $T = 260.6$ to 750.0 K,
– pressure: up to $p = 100$ MPa,
– density: up to $\rho = 1136.5$ kg·m^{-3},
– vapor-liquid saturation line.

The vapor-liquid saturation line defines the limit for extrapolation up to the maximum temperature and the minimum pressure. To calculate the liquid densities with the different models at the saturation line, the corresponding temperatures and pressures from REFPROP Fig. 5 are used. As already shown in the paper from Yang *et al*. [1], the M1.1 Fig. 5 shows the best extrapolation behavior within $\pm 2\,\%$ up to $p = 4.5$ MPa and $T = 655$ K with $\rho \approx 740$ kg·m^{-3}. The graphs for models M1 and M2 in Fig. 5 are similar to each other, while the model M3 shows the largest deviations from the densities calculated with the fundamental equation in REFPROP. It is important to note that none of the models reproduces the curvature of densities calculated from the fundamental equation in REFPROP, which means that they are all strictly limited to the calculation of liquid-phase densities.

Many technical applications (e.g., heating and cooling processes, polymer production and solvation in process technology) work at pressures below 5 MPa. For this reason, we have extrapolated the measured isotherms to ambient pressure. All models reproduce the values obtained by the fundamental equation as available in the current version of REFPROP within relative deviations of $-\,0.28$ to $-\,0.1\,\%$. Each model shows the largest deviation at the 393 K isotherm, and again, the Eureqa model yields the best result with a maximum deviation of $-\,0.20\,\%$. We aim to consider boundary conditions, such as the extrapolation behavior, in further investigations.
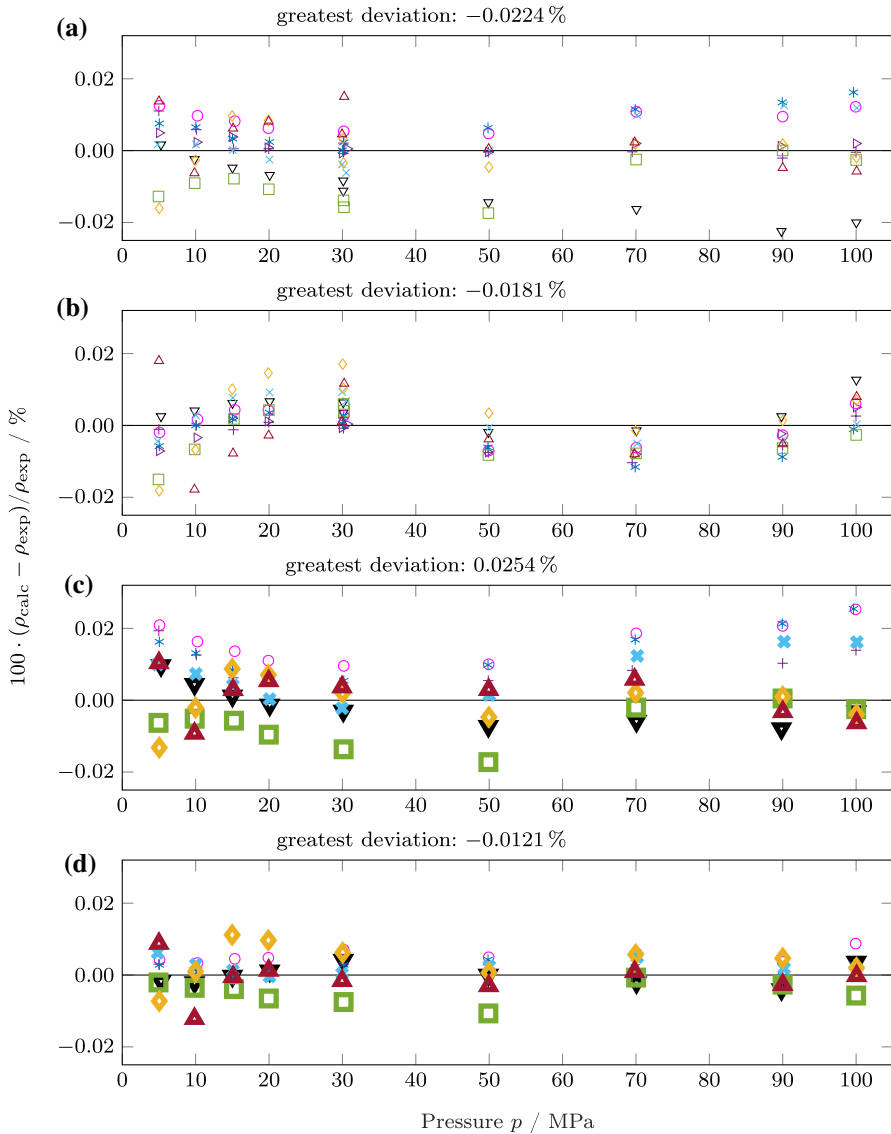
Fig. 3 Relative deviations of **ethylene glycol** densities calculated with the different models in Table 4 from experimental values, (a) M1, (b) M1.1, (c) M2, best five measured isotherms (bold marker), (d) M3, best five measured isotherms (bold marker). $\triangledown$, $T \approx 283.32$ K; $+$, $T \approx 293.142$ K; $\triangleright$, $T \approx 293.140$ K; $\bigcirc$, $T \approx 298.18$ K; $*$, $T \approx 313.13$ K; $\times$, $T \approx 333.09$ K; $\square$, $T \approx 353.13$ K; $\diamond$, $T \approx 373.22$ K; $\triangle$, $T \approx 393.09$ K

5 Conclusions

To decrease the time and even financial expenditure for accurate thermodynamic property modeling, it is necessary to gather suitable experimental data with respect to accuracy and information content. Therefore, we investigated the application of
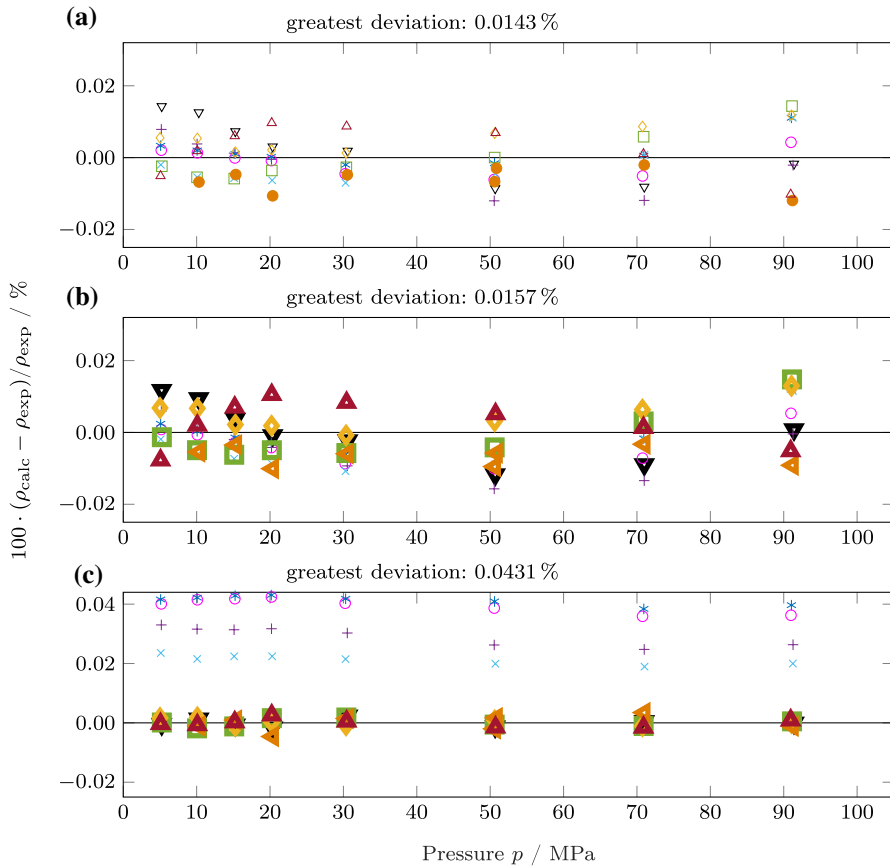
Fig. 4 Relative deviations of **propylene glycol** densities calculated with the different models in Table 4 from experimental values, (a) M1, (b) M2, best five measured isotherms (bold marker), (c) M3, best five measured isotherms (bold marker). $\triangledown$, $T \approx 272.73$ K; $+$, $T \approx 283.18$ K; $\bigcirc$, $T \approx 293.18$ K; $*$, $T \approx 298.12$ K; $\times$, $T \approx 313.12$ K; $\square$, $T \approx 333.05$ K; $\diamond$, $T \approx 352.99$ K; $\triangleleft$, $T \approx 373.34$ K; $\triangle$, $T \approx 392.95$ K

OED to liquid-phase densities of ethylene glycol by calculating the information content of existing experiments along different isotherms. With our resulting model M3 fitted to the most informative five isotherms, we reproduce the measured data with a maximum relative deviation of 0.02542 %. Compared to the maximum relative deviation of the model involving all eight isotherms M1 of 0.02243 %, this is a very promising result. These results are confirmed using the measured data of propylene glycol 0.01571 % compared to 0.01434 %). The smaller deviations for propylene glycol can be explained by the model development originally based on these data. By fitting two new models (M2 for ethylene glycol and propylene glycol) using the best calculated selection of isotherms, we demonstrate that, with OED, sufficiently accurate models can be developed with fewer experiments than traditionally used.

Due to the limitation to already measured data, we are not able to compare these results with a model fitted to the best five freely selected isotherms. At this point, new

Table 5 Parameters of Eqs. 2.1 and 2.2 for the models M1 to M3 in Table 4 for ethylene glycol and propylene glycol

	i	1	2	3	4
Ethylene glycol					
M1	n_i	1.212	-1.159×10^{-1}	-2.295×10^{-3}	1.680×10^{-2}
	t_i	0.0	1.0	4.0	2.5
	p_i	0.0	-0.5	-2.5	-4.0
M1.1	n_i	1.074	2.633×10^{-2}	4.291×10^{-2}	-3.898×10^{-3}
	t_i	0.0	0.0	-1.0	-1.0
	p_i	0.0	1.0	0.0	-1.0
M2	n_i	1.212	-1.159×10^{-1}	-2.532×10^{-3}	1.721×10^{-2}
	t_i	0.0	1.0	4.0	2.5
	p_i	0.0	-0.5	-2.5	-4.0
M3	n_i	1.130	-3.078×10^{-2}	-3.178×10^{-4}	1.007×10^{-2}
	t_i	-4.741×10^{-2}	2.164	3.902	3.519
	p_i	2.173×10^{-2}	-1.048	-2.492	-4.442
Propylene glycol					
M1	n_i	1.140×10^{1}	-1.174×10^{-1}	-2.3745×10^{-3}	1.03×10^{-2}
	t_i	0.0	1.0	4.0	2.5
	p_i	0.0	-0.5	-2.5	-4.0
M2	n_i	1.140	-1.174×10^{-1}	-2.458×10^{-3}	1.104×10^{-2}
	t_i	0.0	1.0	4.0	2.5
	p_i	0.0	-0.5	-2.5	-4.0
M3	n_i	5.796	-4.775	-1.433×10^{-3}	4.104×10^{-3}
	t_i	2.859×10^{-1}	3.739×10^{-1}	3.813	2.257
	p_i	-2.941×10^{-2}	-4.560×10^{-2}	-2.687	-4.140

		5	6	7	8
Ethylene glycol					
M1		-6.42×10^{-3}	3.197×10^{-3}	-5.626×10^{-5}	-1.045×10^{-4}
		3.0	-0.5	3.5	5.5
		-6.0	2.0	2.5	0.5
M1.1		2.540×10^{-3}	-1.008×10^{-2}	-2.544×10^{-2}	-
		2.0	2.0	2.0	-
		1.0	-1.0	0.0	-
M2		-6.458×10^{-3}	3.291×10^{-3}	-8.070×10^{-5}	-6.647×10^{-5}
		3.0	-0.5	3.5	5.5
		-6.0	2.0	2.5	0.5
M3		-3.721×10^{-3}	2.005×10^{-3}	-4.649×10^{-5}	-1.510×10^{-4}
		4.460	-3.598×10^{-1}	3.492	4.829
		-6.173	2.412	2.476	6.854×10^{-1}
Propylene glycol					
M1		-3.074×10^{-3}	3.834×10^{-3}	-2.591×10^{-5}	-1.835×10^{-4}
		3.0	-0.5	3.5	5.5
		-6.0	2.0	2.5	0.5

Table 5 (continued)

	5	6	7	8
M2	-3.480×10^{-3}	3.894×10^{-3}	-1.886×10^{-5}	-1.833×10^{-4}
	3.0	-0.5	3.5	5.5
	-6.0	2.0	2.5	0.5
M3	-1.073×10^{-3}	4.485×10^{-3}	5.140×10^{-3}	-4.790×10^{-4}
	2.783	-6.576×10^{-1}	3.840	6.984
	-6.082	1.646	3.534×10^{-1}	2.583×10^{-1}

measurements become necessary to compare the models with the best five measured and the best five freely selected isotherms, where we are confident to achieve even better results. Finally, the comparison of the model M3, where also the exponents were fitted, shows a significant improvement for ethylene glycol compared to the model M1. However, this is not the case for propylene glycol, where the model M3 shows large deviations regarding the not selected isotherms. This result highlights the importance of the sequential optimal experimental design described at the end of Sect. 1 to select experiments for nonlinear models and the fundamental choice of the functional form. The investigation of the extrapolation behavior underlines this even more, where the use of different thermodynamic criteria beyond the actual measurement data are crucial.

Considering the typical temperature range of thermodynamic measurement devices such as the high-pressure vibrating tube densimeter used in our previous studies, i.e., (273 to 473) K, numerical studies have shown that isotherms near the ambient temperature are often selected last. From an experimental point of view this is advantageous because these are measurements with larger uncertainty since a thermostating task at ambient temperature is often tricky (e.g., a circulating thermostat has to switch between heating and cooling against the ambient temperature).

Our next steps are:

– Measuring the free calculated isotherms, fit the data to the new model and compare the results to the existing models.
– Defining different objective functions in addition to those describing the parameter uncertainty.
– Developing methods to include thermodynamic criteria as boundaries for the model (e.g., extrapolation behavior).
– Using OED based on the nonlinear model (with free exponents) using the sequential process described in Sect. 1.
– Investigating the influence of the number of terms to deal with the problem of overfitting.

With these next steps, we aim to create the basis for an OED setup specialized for the measurement of thermodynamic properties. The data for this paper and the calculations performed are provided in the supporting information.
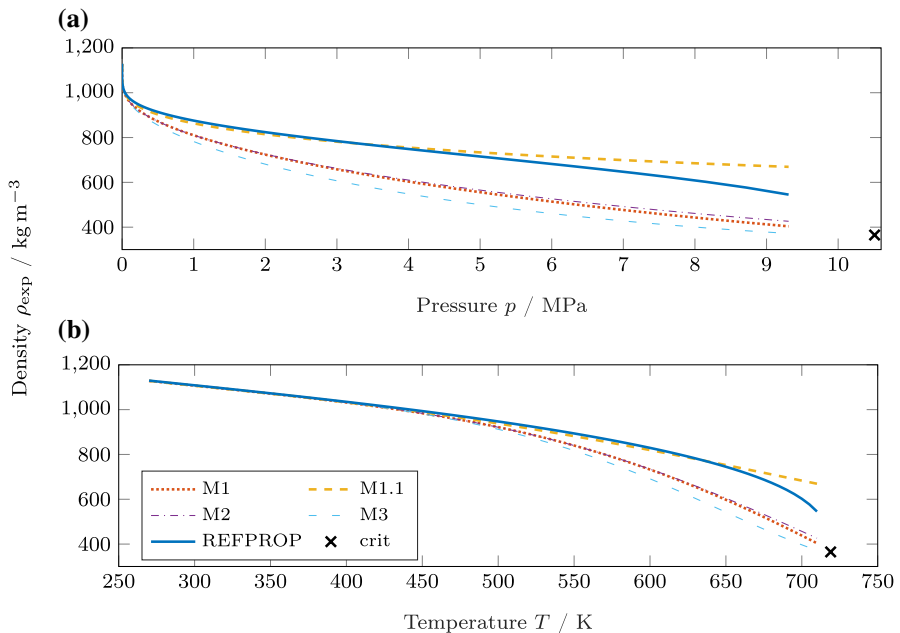
Fig. 5 Saturated liquid densities calculated with the models M1, M1.1, M2 and M3 (see Table 4) for (a) the saturation pressure and (b) the saturation temperature as both computed with the fundamental equation of state as implemented in REFPROP [17]. Saturated liquid densities and the critical point **x** were also calculated with the same fundamental equation of state

References

1. X. Yang, C.C. Sampson, O. Frotscher, M. Richter, J. Chem. Thermodyn. **144**, 106054 (2020). https://doi.org/10.1016/j.jct.2020.106054
2. C.C. Sampson, X. Yang, J. Xu, M. Richter, J. Chem. Thermodyn. **131**, 206 (2019). https://doi.org/10.1016/j.jct.2018.10.016
3. G. Schilling, R. Kleinrahm, W. Wagner, J. Chem. Thermodyn. **40**, 1095 (2008). https://doi.org/10.1016/j.jct.2008.02.020
4. M. Schmidt, H. Lipson, Science **324**, 81 (2009). https://doi.org/10.1126/science.1165893
5. M. Schmidt, H. Lipson. Eureqa (2015). http://formulize.nutonian.com/documentation/eureqa/

6. G. Franceschini, S. Macchietto, Chem. Eng. Sci. **63**, 4846 (2008). https://doi.org/10.1016/j.ces.2007.11.034

7. A. Rasch, H.M. Bücker, A. Bardow, Comput. Chem. Eng. **33**, 838 (2009). https://doi.org/10.1016/j.compchemeng.2008.12.008

8. D. Dechambre, L.W.M. Wolff, C. Pauls, A. Bardow, Ind. Eng. Chem. Res. **53**, 19620 (2014). https://doi.org/10.1021/ie5035573

9. R.E. Gibson, O.H. Loeffler, J. Phys. Chem, **43**, 207 (1939). https://doi.org/10.1021/j150389a003

10. R. Span, *Multiparameter Equations of State* (Springer, Berlin, 2000). https://doi.org/10.1007/978-3-662-04092-8

11. M. Thol, I.H. Bell, in *High-Pressure Flows for Propulsion Applications* (American Institute of Aeronautics and Astronautics, Inc., 2020), pp. 365–407. https://doi.org/10.2514/5.9781624105814.0365.0408

12. U. Setzmann, W. Wagner, Intl. J. Thermophys. **10**, 1103 (1989). https://doi.org/10.1007/bf00500566

13. D. Sommer, R. Kleinrahm, R. Span, W. Wagner, J. Chem. Thermodyn. **43**, 117 (2011). https://doi.org/10.1016/j.jct.2010.08.010

14. A. Pázman, *Foundations of Optimum Experimental Design*, *Mathematics and its Applications*, vol. 14 (Springer, 1986). Translated from the Czech

15. D. Uciński, *Optimal Measurement Methods for Distributed Parameter System Identification*. Systems and Control Series (CRC Press, Boca Raton, FL, 2005). https://doi.org/10.1201/9780203026786

16. A.C. Atkinson, A.N. Donev, R.D. Tobias, *Optimum Experimental Designs, with SAS, Oxford Statistical Science Series*, vol. 34 (Oxford University Press, Oxford, 2007)

17. E.W. Lemmon, H.I. Bell, M.L. Huber, M.O. McLinden, NIST Standard Reference Database 23: NIST Reference Fluid Thermodynamic and Transport Properties Data (REFPROP): Version 10.0 (2018). https://doi.org/10.18434/T4/1502528. https://www.nist.gov/srd/refprop

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.