Journal of Chemical&Engineering Data

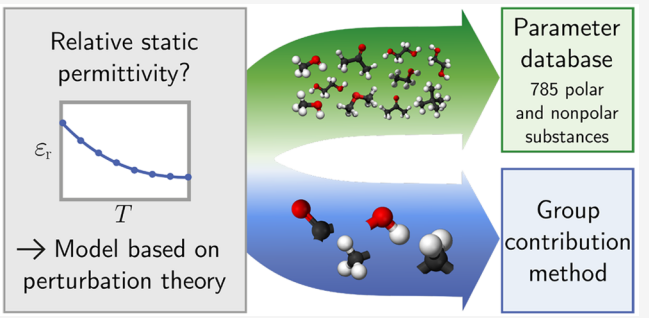
pubs.acs.org/jced
Article

Predicting the Relative Static Permittivity: a Group Contribution Method Based on Perturbation Theory

Lisa Rueben, Johannes Schilling, Philipp Rehner, Simon Müller, Timm Esper, André Bardow,* and Joachim Gross*

Cite This: https://doi.org/10.1021/acs.jced.3c00323

ABSTRACT: The computer-aided design of (bio)chemical processes requires models that predict thermodynamic properties with as little experimental effort as possible. For the important class of electrolyte systems, the relative static permittivity of the solvent is an important thermodynamic property that depends on the temperature, pressure, composition, and molecular structure of the solvent. This work presents a broadly applicable model for the temperature-dependent relative static permittivity of pure and mixed solvents based on perturbation theory, including a group contribution method. For this purpose, we extend our previous model for polar substances to nonpolar substances. The developed model is parametrized for 785 substances, where permittivity and density data are available in the Dortmund Data Bank and the ThermoML database. Subsequently, a group contribution method is developed to predict the permittivity parameters from the molecular structure. With a mean absolute deviation of 0.2 averaged over all 785 substances, the parametrized model accurately correlates the relative static permittivity over a wide range of permittivities and temperatures. Moreover, the group contribution method achieves a mean absolute deviation of 0.6 for the substances in the training set. A leave-one-out cross-validation shows that the group contribution method accurately predicts substances not included in the training set.



1. INTRODUCTION

Electrolyte systems are crucial for decarbonizing energy and process systems, for example, within CO_2 capture processes[1] based on chemical absorption or electrochemical processes.[2,3] An essential task in developing these processes is selecting a suitable pure solvent or solvent mixture. For this purpose, computer-aided solvent selection methods have been developed based on high-throughput screening[4] or computer-aided molecular design[5−7] that evaluate candidate solvents based on their process performance. These solvent selection methods require modeling of the thermodynamic properties of the solvent mixture.

To predict the behavior of electrolyte systems and operating conditions where no or little experimental data are available, thermodynamic electrolyte models such as primitive electrolyte equations of state or quantum-chemistry based g^E models[8,9] have been developed.[10] Primitive electrolyte equations of state regard the solvent as a dielectric continuum with dissolved ions, while the less mature nonprimitive electrolyte equations of state model the interactions between ions and solvent molecules directly. One example of a primitive electrolyte equation of state is the electrolyte PC-SAFT equation of state (ePC-SAFT) developed by the group of Gabriele Sadowski,[11,12] to whom this special issue is dedicated. To model aqueous electrolytes, the PC-SAFT equation of state was extended by a Debye−Hückel term accounting for the long-range interactions due to the presence of ions. Later, the approach was further extended for mixed solvent electrolyte solutions at higher concentrations[13] by including a Born term.[14,15] In primitive electrolyte equations of state, the Debye−Hückel and Born terms require the relative static permittivity of the solvent mixture as input.[16] Similarly, the permittivity is an input to g^E models where accurate modeling of the permittivity is particularly important.[17] The permittivity of the solvent mixture is a function of the temperature, pressure, and mixture composition (T, p, x). Experimental data are scarce for a wide range of operating conditions. Furthermore, to assess solvents for which no experimental permittivity data are available, the permittivity must be predicted from the structure of the solvent molecules.

Special Issue: In Honor of Gabriele Sadowski

Received: May 31, 2023
Accepted: August 1, 2023

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

Currently, electrolyte equations of state often directly use experimental data or empirical expressions for the permittivity.[14,18−20] These empirical expressions are not physics-based and thus require adjusting many parameters to experimental data.[21,22] Further empirical approaches are based on the molecular theory by Kirkwood[23] and Fröhlich[24] to describe the permittivity ε_r as

$$\frac{(\varepsilon_r - 1)(2\varepsilon_r + 1)}{9\varepsilon_r} = \frac{4\pi N_A}{3v}\left(\alpha + \frac{\mu^2 g}{3k_B T}\right) \quad (1)$$

with the Avogadro constant N_A, the molar volume v, the polarizability α, the dipole moment μ, the Boltzmann constant k_B, and the temperature T. The so-called Kirkwood g-factor describes the correlation between neighboring dipoles and thus requires the orientational structure of the liquid.[25] Because the orientational structure is usually not known, empirical approaches adjust the Kirkwood g-factor to experimental data.[25,26] The empirical expressions, for both the permittivity and the Kirkwood g-factor, require extensive experimental data for each compound over a wide range of temperatures and pressures. However, extensive experimental data are often unavailable, particularly in the context of high-throughput screening and computer-aided molecular design.

In theory, the permittivity can be predicted without experimental data if the molecular structure of the fluid is described sufficiently.[27,28] A comprehensive review on molecular theories to predict permittivities is provided by Kournopoulos et al.[27] These molecular theories are currently primarily applied to model fluids, but some can also be applied to real fluids. Like molecular simulation studies,[29] most of the theories require extensive computations and are thus impractical to use within electrolyte equations of state and process design. Maribo-Mogensen et al.[30] predict the Kirkwood g-factor over a wide range of temperatures and pressures by analyzing the molecular orientation and hydrogen-bonding network of associating fluids.

To overcome the computational challenges of evaluating molecular theories for the permittivity, quantitative structure−property relationship (QSPR) methods have been developed that efficiently predict the permittivity from molecular descriptors. To identify a suitable set of molecular descriptors, experimental permittivity data are correlated with various topological, geometric, and electronic descriptors.[31] The molecular descriptors can be obtained from quantum chemistry calculations or databases[31−37] (e.g., highest occupied molecular orbital, dipole moment, van der Waals surface area, or refractive index). Current QSPR methods[31−37] for the permittivity are limited to 25 °C and thus cannot resolve the permittivity's temperature dependence required in process design. In addition, the molecular descriptors might be difficult to obtain in the context of high-throughput screenings or molecular design, particularly if quantum chemistry calculations are necessary.

As an alternative to QSPR methods, the permittivity can be efficiently predicted using group contribution (GC) methods. In general, GC methods allow for the prediction of thermodynamic properties from a molecule's structure, described by its functional groups.[38] In contrast to the quantum-chemical descriptors, the fragmentation into functional groups is easy to obtain. Depending on the variety of functional groups, GC methods can cover a large molecular space. So far, few GC methods have been proposed for the permittivity. Sheldon et al.[39] model the permittivity at 25 °C based on the refractive index for low permittivities and the dipole moment for high permittivities. The refractive index and dipole moment are predicted using a GC method based on the UNIFAC groups. Bouteloup and Mathieu[40] predict the orientational dipolar parameter ($g\mu^2$) within the Kirkwood−Fröhlich equation (cf., eq 1) from the polar groups of the molecule. The orientational dipole parameter ($g_i\mu_i^2$) of group i can be refined depending on the neighboring groups. With the Kirkwood−Fröhlich equation, the model has a physical basis and the set of considered polar groups is comprehensive. However, the temperature dependence of the molar volume is neglected and the model's performance is only considered at near-ambient temperatures. Substances at high or low temperatures are excluded from the training and test set.

Summing up, systematic solvent selection for electrolyte-containing systems requires modeling of the relative static permittivity efficiently for a large set of substances even in the case of scarce experimental data. For this purpose, permittivity models should be leanly parametrized so that the model is either applicable for a large database of candidate fluids or even predictive. To be used in process design and systematic solvent selection, these permittivity models should be easily applicable and quick to evaluate. Moreover, the models should exhibit a physical basis to allow extrapolation in temperature and pressure and possibly also in the substance's molecular structure.

In previous work, we developed a model for the permittivity of pure and mixed solvents based on perturbation theory.[41] For this purpose, we modified the perturbation theory approach by Tani et al.[42] for real fluids and extended it to mixtures. Our model is third order in dipole density and exhibits lean parametrization with only two parameters adjusted to pure component experimental permittivity data. The sound physical basis ensures good extrapolation capabilities even in the case of scarce experimental data: for ethylene glycol as an example, we showed that adjusting one parameter to one data point at 303.15 K and setting the other parameter to the value of water already enables accurately extrapolating permittivities down to 278.15 K and up to 423.15 K. Originally, we applied the model to water, methanol, and ethylene glycol as three polar substances and to the mixtures water/methanol and water/ethylene glycol.[41] Despite the promising results for correlating and extrapolating permittivities, the broad applicability to a larger substance database remains open. In particular, the application to wide ranges of permittivity, from nonpolar substances to very polar ones, has not yet been assessed. Moreover, the study did not ascertain if the model parameters can be predicted from the molecular structure to enable use of the model for solvent design or screening purposes.

In this work, we present a broadly applicable model for relative static permittivity $\varepsilon_r(T, \rho)$ with promising predictive capabilities. First, we extend our previous model[41] to nonpolar substances to expand its applicability (section 2). Then, we investigate the performance of our model over a wide range of permittivities $\varepsilon_r(T, \rho(T))$ by parametrizing 785 substances in the liquid phase, neglecting pressure dependence (sections 3.1 and 4.1). Finally, we develop a group contribution method for the model parameters to predict the permittivity $\varepsilon_r(T, \rho(T, p))$ from the molecular structure of the substance (sections 3.2 and 4.2).

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

2. PERTURBATION THEORY FOR THE RELATIVE STATIC PERMITTIVITY: MODEL EXTENSION TO NONPOLAR SUBSTANCES

In previous work,[41] we modeled the relative static permittivity ε_r for dipolar pure components as

$$\varepsilon_r - 1 = 3y\left[1 + y + \left(\frac{17}{16}I(y) - 1\right)y^2\right] \tag{2}$$

with the dipole density y as

$$y = \frac{4\pi}{9}\beta\rho a_1\mu^2 \tag{3}$$

Here, $\beta = 1/(k_B\,T)$ is the reciprocal of the temperature, μ is the dipole moment, and ρ is the molar density. The correlation integral $I(y)$ is described as

$$I(y) = 1 + a_2(e^{-y} - 1) \tag{4}$$

The model comprises two parameters adjusted to pure component experimental data: the dipole density scaling parameter $a_1 \geq 0$ and the correlation integral parameter $a_2 \in [0,1]$. The dipole density parameter a_1 was introduced to account for the nonsphericality of molecules.

In contrast to polar substances, nonpolar substances do not exhibit a permanent dipole moment. Still, a dipole moment can be induced, depending on the fluid's polarizability α. According to the Clausius−Mosotti relation,[43,44] the dipole density of nonpolar substances is solely a function of density and not of temperature as described by eq 3. Using an effective dipole moment in eq 3, thus, fails to resolve the correct temperature dependence of the permittivity for nonpolar substances. Therefore, we extend our previously developed permittivity model to explicitly account for the polarizability contribution as solely a function of density. For this purpose, we model the effective dipole density y_{eff} as

$$y_{eff} = f(T,\rho) = \frac{4\pi}{9}\rho\Big(\underbrace{\beta a_{11}\mu^2}_{\substack{\text{permanent}\\ \text{-dipole}\\ \text{contribution}}} + \underbrace{3a_{12}\alpha}_{\substack{\text{polarizability}\\ \text{contribution}}}\Big) \tag{5}$$

This definition of the effective dipole density is described by Fröhlich[24] as the sum of the permanent-dipole contribution and the induced dipole due to the polarizability α. Wertheim, in the renormalization theory for polarizable fluids, extends this approach by considering the polarizability as an effective (density- and temperature-dependent) quantity.[45] But this extension is not needed in our model. The resulting effective dipole density y_{eff} shares similarity to the Kirkwood−Fröhlich equation[24] (cf. eq 1)), where parameters a_{11} and a_{12} account for the nonspherical shape of real substances. For nonpolar substances ($\mu = 0$), eq 5 simplifies to

$$y_{eff} = f(\rho) = \frac{4\pi}{3}\rho a_{12}\alpha \tag{6}$$

The extended permittivity model has three parameters: the dipole scaling parameter a_{11}, which corresponds to the parameter a_1 of our earlier work;[41] the polarizability scaling parameter a_{12}; and the correlation integral parameter a_2. In addition to the permanent dipole moment μ, polarizability α is needed. For nonspherical molecules, the polarizability α is a quantity of higher tensorial rank, which also depends on the molecular configuration and on density (due to many-body effects). In our work, we consider the polarizability α to be scalar-valued, defined in a suitable way.[46] We define $a_{11}\mu^2$ and $a_{12}\alpha$ as adjustable parameters characterizing a molecule or a functional group within a molecule. Consequently, the correct definition of the dipole moment μ and of the polarizability α for real molecules is irrelevant. The parameters a_{11} and a_{12} are crucial to describing nonspherical molecules with a perturbation theory derived for spherical species.

As demonstrated in previous work,[41] the developed permittivity model can be applied to mixtures. Thus, we also demonstrate how to apply the extended model (eq 5) to mixtures. For this purpose, the effective dipole density y_{eff} is expressed based on the partial density ρ_i of m components as

$$y_{eff} = \frac{4\pi}{9}\sum_{i=1}^{m}\rho_i[\beta(a_{11}\mu^2)_i + 3(a_{12}\alpha)_i] \tag{7}$$

The correlation integral is mixed based on the mole fraction x_i of component i:

$$I(y_{eff}) = 1 + \left(\sum_{i=1}^{m} a_{2,i}\,x_i\right)\cdot(e^{-y_{eff}} - 1) \tag{8}$$

Optionally, to improve the accuracy, the mixture permittivity model can be extended using a binary parameter ψ_{ij} for the interaction of component i with component j:

$$y_{eff} = \frac{4\pi}{9}\rho\sum_{i=1}^{m}\sum_{j=1}^{m} x_i\,x_j(1 - \psi_{ij})[\beta(a_{11}\mu^2)_{ij} + 3(a_{12}\alpha)_{ij}] \tag{9}$$

with

$$(a_{11}\mu^2)_{ij} = \frac{(a_{11}\mu^2)_i + (a_{11}\mu^2)_j}{2} \tag{10}$$

and

$$(a_{12}\alpha)_{ij} = \frac{(a_{12}\alpha)_i + (a_{12}\alpha)_j}{2} \tag{11}$$

To guarantee consistent pure component permittivities, the binary parameter ψ_{ij} is set to 0 for $i = j$. The binary parameter ψ_{ij} can be adjusted to experimental mixture permittivities.

3. METHOD: PARAMETRIZING THE PERMITTIVITY MODEL BASED ON PERTURBATION THEORY

In the following, we first introduce the methods for the parameter database (section 3.1) and the GC method (section 3.2). The results are presented in sections 4.1 and 4.2, respectively.

3.1. Method: Parametrizing the Permittivity Model for 785 Substances. We parametrize the extended model by minimizing the squared deviations between the experimental data points $\varepsilon_{r,exp,i}$ and the model results $\varepsilon_{r,model,i}$ for all n_p data points. To improve robustness against outliers, we apply a Huber loss function that weights larger residuals $r_i = \varepsilon_{r,model,i} - \varepsilon_{r,exp,i}$ only linearly.[47] Thus, as an objective function, we define the loss $\mathcal{L}_\delta$ as

$$\mathcal{L}_\delta = \sum_{i=1}^{n_p}\mathcal{L}_{\delta,i}(r_i,\,\delta) \tag{12}$$

with the loss $\mathcal{L}_{\delta,i}(r_i, \delta)$ in each data point i as

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

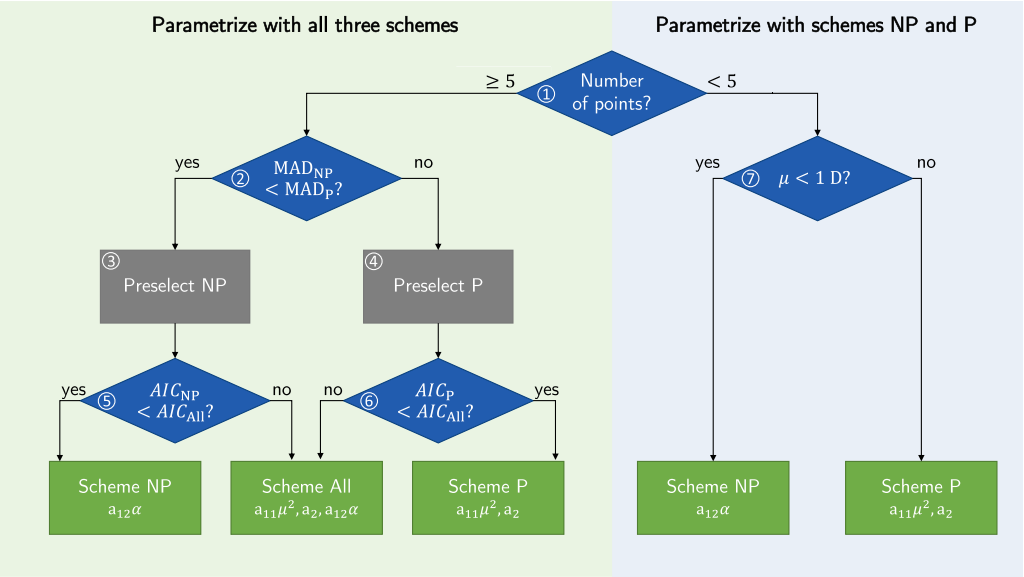


Figure 1. Parametrization procedure for the extended model. For well-measured substances, we parametrize all three schemes and select one of the three schemes with the help of the corrected Akaike information criterion, AIC. For substances with less than five data points, we use information on the dipole moment to choose between scheme NP and scheme P.

$$\mathcal{L}_{\delta,i}(r_i, \delta) = \begin{cases} \frac{1}{2} r_i^2 & \text{for } |r_i| \leq \delta \\ \delta \cdot \left(|r_i| - \frac{1}{2}\delta \right), & \text{otherwise} \end{cases} \tag{13}$$

Residuals below δ are considered in the loss squared, and residuals above δ only linearly. A good choice of δ depends on the order of magnitude of the permittivity. For substances with high permittivities, higher absolute errors are more tolerated than for substances with low permittivities. Therefore, we choose δ for each substance separately as 5% of the mean experimental value for the permittivity. Outliers cannot be defined reliably if only a few data points are available. Therefore, we do not apply the Huber loss function for substances with fewer than five data points.

3.1.1. Parametrization Schemes. To parametrize our model for a pure component, we apply one of three parametrization schemes:

- **NP**: Nonpolar substances are described by the scheme NP. For these substances with low permittivities, the effective dipole y_{eff} (eq 5) is polarizability-dominated. Consequently, we adjust the polarizability term $a_{12}\,\alpha$ to experimental data, while the permanent-dipole term $a_{11}\,\mu^2$ is set to zero. Generally, one could additionally adjust the correlation integral parameter a_2. However, the model is insensitive to the correlation integral at low permittivities. Therefore, we set the correlation integral parameter a_2 to zero. The scheme NP results in a one-parameter model.
- **P**: Polar substances are described by the scheme P. For these substances, the effective dipole y_{eff} (eq 5) is dominated by the permanent dipole. Thus, we adjust the permanent-dipole term $a_{11}\mu^2$ and the correlation integral parameter a_2 to experimental data, while the polarizability term is set to zero. The scheme P results in a two-parameter model.
- **All**: For the substances parametrized with the scheme All, both the permanent and the polarizability-induced dipole moment contribute to the effective dipole moment. The results can be improved significantly upon adjusting all three parameters ($a_{11}\mu^2$, $a_{12}\alpha$, and a_2) to experimental data. Consequently, scheme All results in a three-parameter model.

For each substance of our data set, we select one of the three parametrization schemes according to the parametrization procedure given in boxes 1 to 7 in Figure 1. Generally, we assume a minimum of five data points necessary to parametrize a substance with all three schemes (left side of Figure 1). Although adjusting the full set of parameters will reduce the mean deviation, we prefer the parametrization scheme with fewer parameters if the benefit of adding further parameters is small. Therefore, we preselect either scheme NP or P based on the lower mean absolute deviation (MAD; boxes 2−4 in Figure 1). Then, we assess whether the error reduction achieved by the scheme All justifies the additional parameters. For this purpose, we compare the preselected scheme NP or P with scheme All using the corrected Akaike information criterion, *AIC*[48] (boxes 5 and 6 in Figure 1). In contrast to the Bayesian information criterion, the corrected Akaike information criterion is also valid for small sample sizes with a number of data points to number of parameters ratio n_p/K smaller than 40.[49] The corrected Akaike information criterion, *AIC*, is calculated as[49]

$$AIC = n_p \cdot \ln \sigma^2 + 2K + \frac{2K \cdot (K+1)}{n_p - K - 1} \tag{14}$$

with the variance σ^2, the number of data points n_p, and the total number of adjustable parameters K. In the least-squares case, the variance σ^2 can be computed from the sum of the residuals as[49]

$$\sigma^2 = \frac{\sum_{i=1}^{n_p} (\varepsilon_{r,model,i} - \varepsilon_{r,exp,i})^2}{n_p} \tag{15}$$

with the model results $\varepsilon_{r,model,i}$ and the experimental data points $\varepsilon_{r,exp,i}$. The corrected Akaike information criterion weights the gain in the fit versus the number of parameters K employed. Comparing the preselected scheme (NP or P)

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

and All, we choose the scheme that minimizes the corrected Akaike information criterion.

If less than five data points are available for a substance, we enforce either scheme NP or P to avoid overfitting with scheme All (right side of Figure 1). We do not decide between the two schemes using mean deviations because the mean deviations do not reliably discriminate between schemes for substances with few data points, particularly if the data points are close regarding temperature. Instead, we use information on the polarity, enforcing the more physical model. As a metric for the polarity, we use the dipole moment obtained from the DIPPR database.[50] If no dipole moment is available in DIPPR, the dipole moment is calculated by ORCA[51,52] after a geometry optimization of each structure. If the dipole moment is greater than or equal to 1 D, we choose scheme P. If the dipole moment is lower than 1 D, we choose scheme NP. We determine this threshold value of μ = 1 D by analyzing the dipole moments of all well-measured substances with five or more data points parametrized with schemes NP and P (cf., Supporting Information, section S2.1). If only one data point is available, we fix the correlation integral parameter a_2 to the value of water (a_2 = 0.12) for the scheme P to obtain a one-parameter model, as shown in Neumaier et al.[41]

3.1.2. Liquid-Density Correlation. In addition to the model parameters, the permittivity model requires density ρ as an input (eq 5). If employed within the context of equations of state, the density can be calculated with the respective equation of state. However, we aim to establish a permittivity model with a permittivity parameter database independent of a specific equation of state. Therefore, we provide a simple correlation for the liquid density $\rho(T)$ that can be easily and efficiently used with the provided permittivity parameter database. Generally, the liquid density, $\rho(T, p)$ is a function of temperature and pressure. However, most data points for the liquid density and the permittivity have been collected at ambient or saturation pressure, and often even no pressure is specified with the experimental data. Thus, we neglect the pressure dependence of the liquid density $\rho(T)$ resulting in a correlation for the experimental densities at pressure $p(T)$ = max[1 bar, $p_{sat}(T)$].

We parametrize the liquid-density correlation provided by Yaws and Pike:[53]

$$\rho(T) = \rho_{crit} A^{-\left(1-\frac{T}{T_{crit}}\right)^B} \tag{16}$$

with the critical density ρ_{crit}, the critical temperature T_{crit}, and two adjustable parameters A and B. If available, we use estimates for the critical temperature and density that Bell et al.[54] collected from various sources, such as the NIST webbook or the CRC handbook. Details on the liquid-density parametrization are provided in the Supporting Information, section S2.2. With the liquid density correlation, our model provides a simple yet physical approach to the permittivities of a large set of substances.

To summarize, our model $\varepsilon_r(T)$ is constituted by eqs 2, 4, 5, and 16 with up to three parameters for the permittivity model (the dipole scaling parameter $a_{11}\mu^2$, the polarizability scaling parameter $a_{12}\alpha$, and the correlation integral parameter a_2) and four parameters for the density correlation (A, B, T_{crit}, and ρ_{crit}) adjusted independently using density data.

3.1.3. Experimental Data. We parametrize our model for all substances where at least one experimental data point for the permittivity and at least four data points for the liquid density are available in the Dortmund Data Bank[55] and the ThermoML database.[56,57] The available experimental permittivity data are prefiltered based on phase, pressure, and frequency. We only include liquid phase data, because the density correlation was developed for the liquid phase. Furthermore, the pressure is restricted below 5 bar to be able to assume that the density is independent of pressure. Finally, only relative static permittivities are included. For this purpose, only measurements with frequencies below 1000 Hz are considered, which are frequencies sufficiently close to zero. As a result, our model is parametrized for 785 substances covering a wide range of permittivities from 1 to 350. More information on the experimental data such as the distribution of temperatures or permittivities is available in the Supporting Information, section S1.

3.2. Method: Predicting the Permittivity Using a Group Contribution Method. Next to parametrizing the model individually for 785 substances, we develop a GC method for predicting the permittivity based on the molecular structure. For this purpose, we choose a group contribution approach that models the molecule as a combination of functional groups. The molecule is fragmented into a set of functional groups $\mathcal{G}$, with n_i indicating the number of occurrences of group i in the molecule. For example, 1-propanol consists of one CH_3 group, two CH_2 groups, and one OH group.

We define a GC approach to calculate the permittivity parameters $a_{11}\mu^2$, $a_{12}\alpha$, and a_2, applying an additive group contribution ansatz:

$$a_{11}\mu^2 = \sum_{i\in\mathcal{G}} n_i (a_{11}\mu^2)_i \tag{17}$$

$$a_{12}\alpha = \sum_{i\in\mathcal{G}} n_i (a_{12}\alpha)_i \tag{18}$$

$$a_2 = \sum_{i\in\mathcal{G}} n_i a_{2,i} \tag{19}$$

Here, $(a_{11}\mu^2)_i \geq 0$, $(a_{12}\alpha)_i \geq 0$, and $a_{2,i} \in [0,1]$ denote the parameters of the functional group i. The structure of the permittivity model, eqs 2, 4, and 5, remains unaltered.

The densities $\rho(T, p)$ are calculated using the PC-SAFT equation of state[58] with the homosegmented GC method for the PC-SAFT parameters.[59] Sauer et al.[59] parametrized the functional group parameters within the homosegmented GC method to saturation pressures and liquid densities for alkanes, alkenes, 1-alkynes, cyclopentanes, cyclohexanes, alkylbenzenes, aldehydes, ketones, ethers, formates, esters, 1-alkanols, and 1-amines. For all molecular families studied by Sauer et al.,[59] the mean absolute relative deviation (MARD) for the liquid densities is below 7%. Thus, the GC method for permittivity developed in this work allows for the calculation of both permittivity and other thermodynamic equilibrium properties in a unified, consistent framework.

Consequently, we adopt the set of functional groups parametrized by Sauer et al.[59] for the permittivity. These functional groups are nonpolar and polar groups with carbon, nitrogen, and oxygen (Table S2 in Supporting Information, section S3.1). We limit our consideration to substances with one polar or associating group because the homosegmented GC method was only parametrized and tested for this case: dipole moments do not satisfy group addivity due to potential symmetry effects that lead to dipole-moment cancellation.[60]

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

The homosegmented GC method omits the 3D structure of the molecule and thus lacks information to construct a proper molecular dipole moment for more than one polar group.

For parametrization, we reduce our experimental data set to substances that can be represented by the available functional groups. Functional groups are included in the parametrization only if they occur within at least two substances: Of the 785 substances in the database (cf., section 3.1.3), 220 substances can be represented using the set of functional groups, while 36 new substances can be represented that are not part of the database due to a lack of density data. As a result, we include 256 substances in our group contribution parametrization. The included substances are listed in the Supporting Information, section S3.2.

In contrast to the simplified density model in section 3.1, the densities $\rho(T, p)$ calculated by PC-SAFT depend on the pressure. If no pressure is indicated for an experimental data point, then we assume $p(T) = \max[1\ \text{bar}, p_{\text{sat}}(T)]$.

The parameters for all functional groups are adjusted simultaneously. For the nonpolar and nonassociating groups except the alkynyl group, we adjust the polarizability term with $(a_{12}\alpha)_i \geq 0$, while the other parameters are set to zero. For the polar and associating groups and the alkynyl group, we adjust the permanent-dipole term with $(a_{11}\mu^2)_i \geq 0$ and the correlation integral parameter with $0 \leq a_{2,i} \leq 1$. For these functional groups, the polarizability parameters $(a_{12}\alpha)_i$ are set to zero because the model is insensitive to the polarizability parameters $(a_{12}\alpha)_i$. We treat the alkynyl group as a polar group, because alkynes larger than ethyne are slightly polar,[61] caused by the different hybridization of the single- and the triple-bonded carbon.[60] Moreover, the results for the alkynes improve significantly if they are modeled as polar. Consequently, nonpolar substances are described with one parameter $(a_{12}\alpha)$ and polar or associating substances with all three parameters.

We parametrize the group contribution model by minimizing the squared absolute deviation between the experimental data points $\varepsilon_{r,\text{exp},s,i}$ and the model results $\varepsilon_{r,\text{model},s,i}$ for all substances s. We weight the sum of these residuals for each substance s from the set of substances $\mathcal{S}$ with the square root of the number of points $n_{p,s}$. Thus, the loss $\mathcal{L}$ is calculated as

$$\mathcal{L} = \sum_{s\in\mathcal{S}} \frac{\sum_{i=1}^{n_{p,s}} (\varepsilon_{r,\text{model},s,i} - \varepsilon_{r,\text{exp},s,i})^2}{\sqrt{n_{p,s}}} \tag{20}$$

The number of points is distributed unevenly in the data set; i.e., a few substances comprise numerous points, while most substances are described through 10 points or less. If all points were weighted equally, substances with numerous points would dominate the fit. In contrast, if all substances were weighted equally, substances with few data points would be considered similarly important to well-measured substances. As a compromise between particularly high weights for the well-measured substances and equal weights for all substances, we weight the substances according to the square root of the number of data points. A comparison between various weighting methods is provided in the Supporting Information, section S3.3.

In contrast to the parametrization described in section 3.1, we do not apply a Huber loss function to the GC parametrization. The Huber loss function is suited for coping with outliers in a problem with otherwise Gauss-distributed deviations. By contrast, in a GC approach, we expect some systematic deviations, for which a Huber loss function is inappropriate.

For correlation integral parameters a_2 above 1/17, the model can produce unphysical maxima of the permittivity at very low temperatures with high densities, as described by Neumaier et al.[41] For the three substances previously studied,[41] these unphysical maxima occurred at densities far higher than the experimental densities. However, there is no guarantee for all substances that the respective densities are unphysically high, particularly for small, high-density substances. As an example, for 2-propanone, unphysical maxima are observed at realistic densities. To guarantee that no unphysical maxima occur for the substances in the GC fit, we constrain the derivative of the permittivity over the temperature at the lowest temperature of the corresponding experimental data to $(d\varepsilon_r/dT)_p \leq 0$. Additionally, we report the minimum temperature at which the model can be applied in the Supporting Information for each substance in the GC fit. For user-defined substances not contained in the GC fit, we provide a Python tool to compute the minimum temperature corresponding to the maximum density where the permittivity model can be used.

4. RESULTS

First, section 4.1 presents the results for the parametrization of the permittivity model for 785 substances to establish the permittivity parameter database, as described in section 3.1. Subsequently, section 4.2 analyzes the performance of the developed GC method.

4.1. Results: Parametrizing the Permittivity Model for 785 Substances. The established parameter database developed, according to the parametrization described in section 3.1, contains 785 substances with permittivities between 1 and 350, and temperatures between 1 and 560 K with a mean temperature of 300 K. For 315 substances, less than five data points are available, so these substances are parametrized using information on the dipole moment. Of the 785 substances, 147 substances are classified as nonpolar substances with the effective dipole being polarizability dominated (scheme NP). With 467 substances, the majority of substances are classified as polar with the effective dipole being dominated by the permanent dipole (scheme P). For 171 substances, the error reduction justifies adjusting all three model parameters according to scheme All (Table 1). All substances are listed in the data file in the Supporting Information with their respective scheme, their parameters, and the absolute and relative errors.

We measure the absolute error with the mean absolute deviation (MAD) between the model results $\varepsilon_{r,\text{model},i}$ and the experimental data points $\varepsilon_{r,\text{exp},i}$, as

$$\text{MAD} = \frac{1}{n_{p,s}} \sum_{i=1}^{n_{p,s}} |\varepsilon_{r,\text{model},i} - \varepsilon_{r,\text{exp},i}| \tag{21}$$

The relative error is assessed with the mean absolute relative deviation (MARD) as

$$\text{MARD} = \frac{1}{n_{p,s}} \sum_{i=1}^{n_{p,s}} \left| \frac{\varepsilon_{r,\text{model},i} - \varepsilon_{r,\text{exp},i}}{\varepsilon_{r,\text{exp},i}} \right| \tag{22}$$

With a mean MAD of 0.2, a standard deviation σ_{MAD} of 0.59, and a mean MARD of 1.6%, the model performs accurately on average: with 744 of 784 substances, almost 95% of the

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

Table 1. Statistics of the 785 Substances Parametrized in This Work Using the Mean Absolute Deviation (MAD) and the Mean Absolute Relative Deviation (MARD): Distribution of Parametrization Schemes NP (Nonpolar Substances), P (Polar Substances), and All (All Three Parameters Adjusted)[a]

scheme	# substances	$\overline{\text{MAD}}$	σ_{MAD}	MAD < $\overline{\text{MAD}}$ + σ_{MAD}	$\overline{\text{MARD}}$/%
NP	147	0.03	0.07	135	1.2
P	467	0.29	0.75	442	2.0
All	171	0.09	0.18	153	0.7
	785	0.2	0.6	747	1.6

[a]For all schemes, the mean MAD ($\overline{\text{MAD}}$), standard deviation of MAD (σ_{MAD}), number of substances with a MAD lower than $\overline{\text{MAD}}$, and the mean MARD ($\overline{\text{MARD}}$) are listed.

substances achieve an MAD lower than $\overline{\text{MAD}}$ + σ_{MAD} = 0.79 (Table 1).

Figure 2 shows the permittivity over the temperature for exemplary substances in various permittivity ranges. Cyclohexane is a low-permittivity substance parametrized with scheme NP, while 1-propanol and chloromethane are polar substances parametrized with scheme P and scheme All, respectively. With permittivities up to 300, n-methylformamide is a highly polar substance and can be described by scheme P. The corresponding figures for all substances are provided in the Gitlab Repository published with this paper.

For low-permittivity substances parametrized with scheme NP, the mean MAD is 0.03 with a standard deviation of 0.07. The mean MARD is 1.1%, and 137 of 147 substances exhibit an MAD below 0.1 (Table 1). For two substances, the MARD exceeds 10%. However, these two substances have only two or three contradicting data points from different data sources.

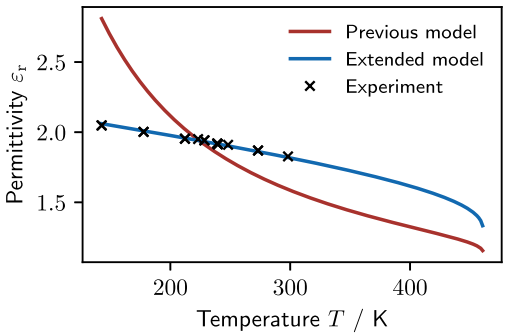


Figure 3. Modeled permittivity vs experimental data over temperature for 2-methylbutane without (red, previous model[41]) and with (blue, extended model) consideration of polarizability.

Thus, the model result is a compromise between contradicting data sources. Moreover, at low permittivities, already relatively low absolute errors lead to high relative errors. To illustrate this result, we show the permittivity over the temperature for these two substances in the Supporting Information, section S2.4.

For substances with low permittivities, the model extension (eq 5) now allows resolving the correct temperature dependence of the permittivity. As an example, Figure 3 compares the model results from the previously developed model[41] (corresponding to scheme P) with the model extension (scheme NP) for 2-methylbutane. Applying the extended model with scheme NP instead of the previous model[41] reduces the MARD from 10.0% to 0.3% by capturing the correct temperature dependence. Thus, the model extension of this work describes the permittivity for nonpolar substances accurately where the effective dipole is polarizability dominated.

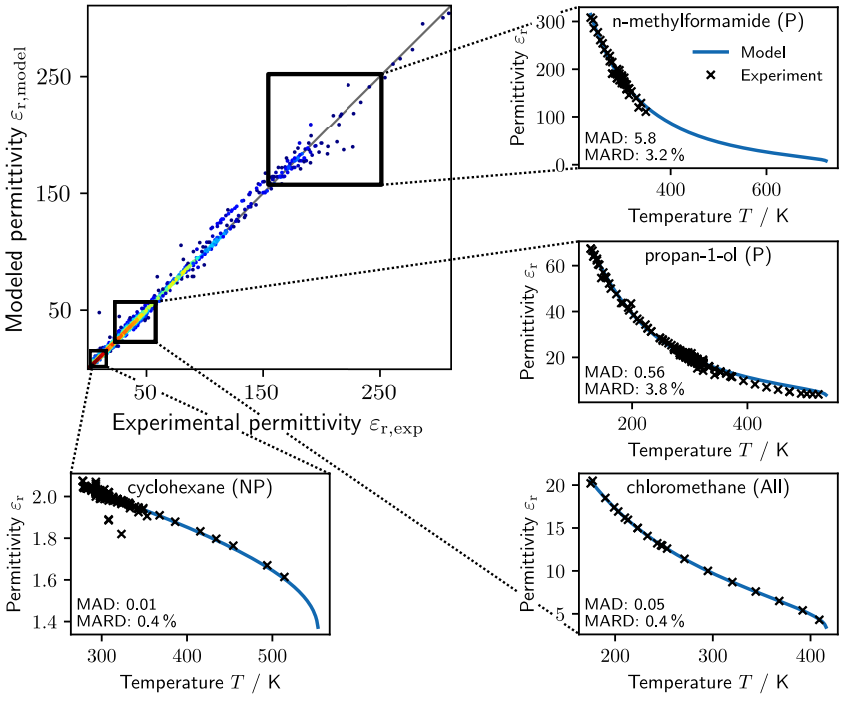


Figure 2. Modeled permittivity vs experimental data for all data points under consideration. Moreover, the modeled permittivities and experimental data are shown over the temperature for four exemplary substances: cyclohexane (parametrized with scheme NP), chloromethane (parametrized with scheme All), 1-propanol, and n-methylformamide (both parametrized with scheme P).

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

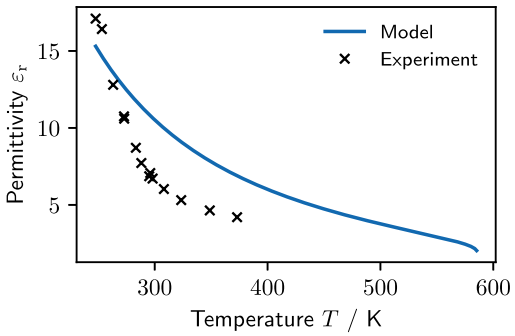


Figure 4. Modeled permittivity vs experimental data over temperature for heptan-3-ol.

The majority of substances are parametrized with scheme P. The errors for scheme P are again small but slightly higher than for the other schemes: the mean MAD is 0.29 with a standard deviation of 0.74, and the mean MARD is 2.0%. Still, 95% of all substances (443 of 467) exhibit an MAD below $\overline{\text{MAD}} + \sigma_{\text{MAD}} = 1.03$. We analyze the model permittivities depending on the temperature for all substances to determine the sources of the higher errors. First, no outlier detection is performed in this work. Besides measurement errors, sources for outliers may include reporting errors, e.g., for the frequency or lack of reported pressure. While the Huber loss function improves the robustness of our parameter fit against outliers, outliers are still included in the reported errors.

Higher errors also occur for polar substances with comparably low permittivities ($\varepsilon_r < 15$) where the permittivity sharply decreases with increasing temperature. These substances are predominantly heavy alcohols (for a detailed list, see Supporting Information, section S2.5). At low permittivities, the two-parameter model is insensitive toward the correlation integral parameter a_2, because the product of the correlation integral and the cubic dipole density is small. Consequently, the two-parameter model reduces to a one-parameter model with only the dipole scaling parameter $a_{11}\mu^2$. This one-parameter model cannot resolve the sharp decrease in permittivity over temperature as shown, e.g., in Figure 4 for heptan-3-ol. For these substances, the correlation integral parameter a_2 becomes 0. If the lower bound ($a_2 \geq 0$) of the correlation integral parameter a_2 is relaxed, then the steep decrease in permittivity can be resolved. However, high negative values are thereby observed for a_2 due to the insensitivity toward a_2. We therefore decide to keep the lower bound of the correlation integral parameter $a_2 \geq 0$ active, as this issue arises for just a few substances. Even for these substances, the model still shows good qualitative agreement.

The substances parametrized with all three model parameters (scheme All) are well-described with a mean MAD of 0.1 and a standard deviation of 0.21, a mean MARD of 0.8%, and 151 of 171 substances with an MAD below $\overline{\text{MAD}} + \sigma_{\text{MAD}} = 0.31$. For all substances, the MARD is below 10%. For only two substances, the MAD exceeds 1 due to outlying experimental data.

Overall, the developed model produces accurate results over a wide range of permittivities from 1 to 300 and temperatures from 1 to 560 K. Furthermore, the model extension developed in this work also enables accurate modeling of the permittivity for nonpolar substances where the effective dipole is determined by polarizability effects.

Mixture Permittivities. The developed permittivity model can be applied to mixtures. In previous work,[41] we assessed the model performance for two mixtures of polar substances: water/methanol and water/ethylene glycol. The model extension presented in section 2 of this work enables us to predict mixture permittivities also for mixtures of polar and nonpolar substances like the binary mixture 1,4-dioxane (nonpolar)/1,2-butanediol (polar) shown in Figure 5 (left side) for a binary parameter of $\psi_{ij} = 0$.

For highly nonideal mixtures that cannot be accurately described with $\psi_{ij} = 0$, we can capture the nonideal behavior by adjusting ψ_{ij} to experimental data, like for the binary mixture glycerol/*n*,*n*-dimethylformamide (Figure 5, right side).

Overall, the extended permittivity model enables the calculation of mixture permittivities also of nonpolar with polar substances. A systematic assessment of the model performance for a larger set of mixtures should be performed in future work.

4.2. Results: Predicting the Permittivity Using a Group Contribution Method. In the following section, we first analyze the correlation results for the developed GC method, i.e., how well the GC method performs on the training data (section 4.2.1). Then, we perform a leave-one-out (LOO) cross-validation to assess how accurately the GC method predicts permittivities of unknown substances (section 4.2.2).

4.2.1. Correlation results for the group contribution method. To analyze the correlation results of the developed GC method, we compare the GC results on the training data to the results from individual parametrization using the PC-SAFT equation of state as a thermodynamic model to calculate the densities (Figures 6 and 7). The PC-SAFT parameters are

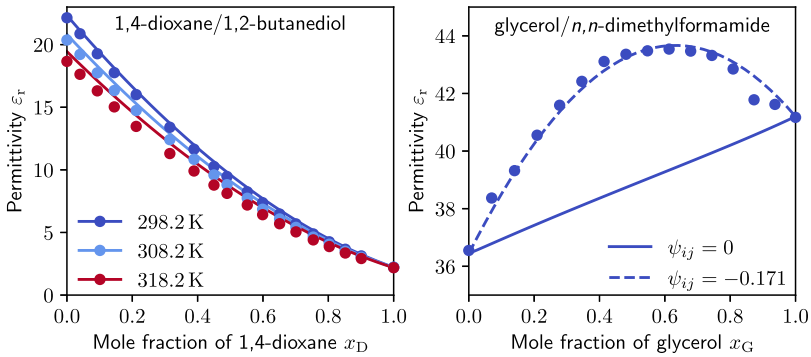


Figure 5. Mixture permittivity over mole fraction for the binary mixture 1,4-dioxane/1,2-butanediol for three temperatures (left side) and the binary mixture glycerol/*n*,*n*-dimethylformamide for $T = 303.15$ K (right side).

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

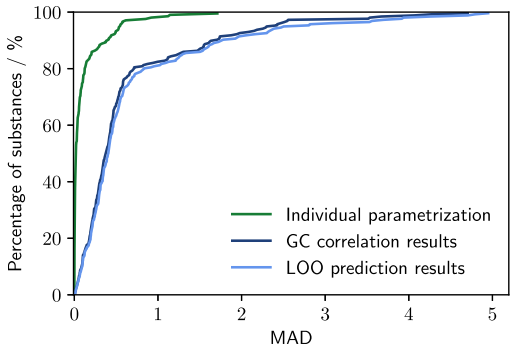


Figure 6. Cumulative distribution of the MAD for the individual parametrization using PC-SAFT, the correlation results, and the prediction results obtained by leave-one-out cross-validation (Section 4.2.2).

taken from Esper et al.[62] For the individual parametrization as a benchmark, only 206 substances are included, because PC-SAFT parameters are not available for all 256 substances. Apart from errors introduced by the GC method for the permittivity model, the total error also contains inaccuracies from the density prediction with the homosegmented GC method for PC-SAFT. The GC method achieves a mean MAD of 0.62 with a standard deviation of 0.78. 88% of the substances achieve an MAD lower than $\overline{\mathrm{MAD}} + \sigma_{\mathrm{MAD}} = 1.4$. Compared to the individual parametrization (mean MAD of 0.10 and standard deviation of 0.19), the GC method is worse but still shows accurate results, given the substantially fewer parameters of the model (27 in total instead of 360).

With an MAD of 4.7, the substance 7,7-dimethyloctan-1-ol exhibits the highest error of all substances. 7,7-dimethyloctan-1-ol is one of five branched alcohols with a >C< group included in the training data. All five of these branched alcohols exhibit MADs above 2. Similarly, the model also performs worse for branched alcohols with >CH group than for linear alcohols. However, for these branched alcohols, the quantitative agreement is adequate and the qualitative agreement good (see respective permittivity-temperature diagrams in the Gitlab repository).

Substances with rare group combinations enter the GC fit with a low weight, which leads to less accurate results. For example, the two substances with the second and third highest MAD are 3-methylbut-3-en-2-one as the only ketone with an alkene group and 3-phenylpropan-1-ol as the only aromatic with a hydroxyl group. The results for the substances with rare group combinations can be improved with more experimental data. However, with MADs around 4, even the substances with the highest errors are described well.

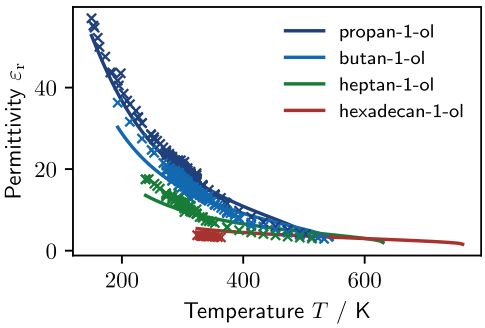


Figure 8. Modeled permittivity and experimental data for selected n-alcohols over temperature using the GC method.

We analyzed the average MAD per molecular family. In Figure 7, only substances are shown where PC-SAFT parameters are available for the individual parametrization. For most families, the GC fit leads to excellent results, with an average MAD below 0.5: All nonpolar families and the polar families of the ethers, esters, and amines exhibit such low errors. Expectedly, the MAD is higher for families that cover a wider range of permittivities, e.g., the n-alcohols and ketones. For n-alcohols (MAD = 1.4), the parametrization of the hydroxyl group is a compromise between 38 substances of high and low permittivity, as the maximum permittivity ranges from 3 to 63. Moreover, already the individual parametrization for n-alcohols leads to an average MAD of 0.4 (cf., Figure 4). As an example, Figure 8 shows selected n-alcohols of high and low permittivity. Despite the higher errors, the model agrees qualitatively well with the permittivity data for all substances. The permittivity is underestimated for propanol, butanol, and heptanol only at low temperatures, where few experimental data are available. Most of the data points are at around room temperature. Therefore, the weight of the low temperatures within the GC fit is particularly low.

For ketones, excellent results are obtained for the permittivity over a wide range of permittivities, as shown in Figure 9 for the selected ketones. The unphysical behavior of

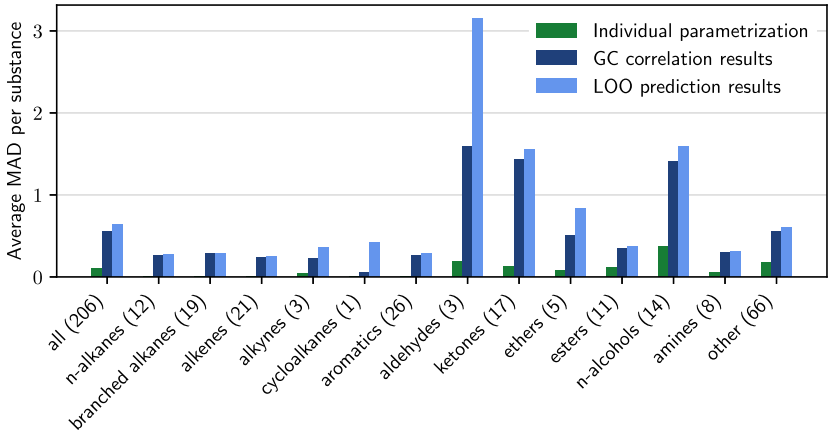


Figure 7. MAD for considered molecular families for the individual parametrization using PC-SAFT, the GC fit, and the leave-one-out cross-validation (Section 4.2.2). The number behind the family indicates the number of substances considered for this family.

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

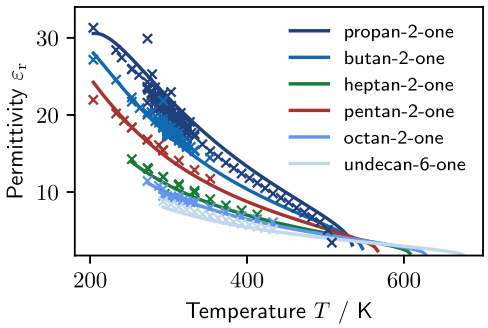


Figure 9. Modeled permittivity and experimental data for selected ketones over temperature using the GC method.

substances with a high density at low temperatures discussed in section 3.2 is observed for propan-2-one as a fluid with a high density at low temperatures. Consequently, our model should not be used for propan-2-one below 200 K, which is the lowest temperature for which experimental data are available.

We compare our GC results to the GC method developed by Bouteloup and Mathieu[40] (cf., section 1) on the data set of our GC method. With 74 groups, Bouteloup and Mathieu[40] parametrize substantially more groups compared to our model (19 groups) and are thus able to represent more substances. However, their GC method neglects the temperature dependence of the density, and the authors conduct their analysis mostly at around room temperature. The average errors per substance of our proposed GC method (average MAD per substance of 0.62) are comparable with the method by Bouteloup and Mathieu[40] (average MAD per substance of 0.61). This result can be attributed to the high share of data points at around room temperature in the data set, for which the model by Bouteloup and Mathieu performs particularly well. In contrast, our model also performs well at high and low temperatures. Although only a few data points at high and low temperatures are available for the training, our model achieves similar errors over the entire temperature range (Figure S6, Supporting Information).

Overall, the developed GC method for the relative static permittivity shows good results for correlating permittivities. For all substances, the permittivity's order of magnitude is captured. In view of only 27 adjustable parameters, the results are convincing, even for the molecular families that cover a wide range of permittivities. The functional group parameters are provided in Table 2.

4.2.2. Predictive Capabilities of the Group Contribution Method. To assess the predictive capabilities of the developed GC method, we perform a leave-one-out (LOO) cross-validation. For this purpose, the GC fit is carried out 256 times. Each time, the data for one test substance is removed from the training data. The parameters obtained from the GC fit are used to predict the permittivity of the test substance and to calculate the errors of the prediction.

The LOO cross-validation achieves a mean MAD of 0.69 with a standard deviation of 0.89. 88% of the substances achieve an MAD lower than $\overline{\mathrm{MAD}} + \sigma_{\mathrm{MAD}} = 1.58$. Compared with the correlation results of the GC fit, the MAD only increases slightly (Figure 7). For most substances and molecular families, the correlation and prediction errors are similar (Figures 6 and 7). The similar errors of correlation and prediction demonstrate the promising extrapolation capabilities of the developed GC method. Expectedly, the results for

Table 2. Group Contribution Parameters for the Functional Groups

group	$(a_{11}\mu^2)_i$/D^2	$a_{2,i}$	$(a_{12}\alpha)_i$/Å^3
$-CH_3$	0.0	0.0	0.0
$-CH_2-$	0.0	0.0	3.364
>CH−	0.0	0.0	6.581
>C<	0.0	0.0	1.476
$=CH_2$	0.0	0.0	0.0
=CH−	0.0	0.0	3.028
>C=	0.0	0.0	9.402
$>C^{Arom}=$	0.0	0.0	4.125
$-CH^{Arom}=$	0.0	0.0	2.707
$>CH^{Hex}-$	0.0	0.0	4.177
$-CH_2{}^{Hex}-$	0.0	0.0	1.818
−C≡CH	0.942	1.000	0.0
−CH=O	7.345	0.0	0.0
>C=O	9.193	0.2151	0.0
$-O-CH_3$	2.507	0.4040	0.0
$-O-CH_2$	2.618	0.0983	0.0
−O−(C=O)−	4.233	0.4821	0.0
−OH	7.582	0.1557	0.0
$-NH_2$	2.258	0.0240	0.0

the substances with rare functional groups are less robust to prediction, particularly when the permittivities vary substantially among those substances. For example, the three aldehydes benzaldehyde (16 data points), butyraldehyde (2 data points), and heptanaldehyde (1 data point) are the only substances in the training set with an aldehyde group. Here, benzaldehyde is the aldehyde with the highest permittivity between 15 and 20, while butyraldehyde exhibits permittivities between 10 and 15, and heptanaldehyde is below 10. For these aldehydes, the average MAD per substance is the highest of all the molecular families.

Typically, the predictive capabilities of GC methods are limited for rare combinations of functional groups (cf., section 4.2.1). As a reference for the user wanting to apply the model to extrapolate, we provide a list of the substances included in the GC fit in the Supporting Information, section S3.2. An overview of the included functional group combinations is provided in Figure S4 in Supporting Information section S3.1. This overview can be consulted to estimate the prediction uncertainty qualitatively. If the group combination has not been included in the training set for the developed GC method, the predictions should be treated with more caution than if the combination occurs several times.

The robustness of the extrapolation is evaluated by analyzing the variability of the obtained functional group parameters $(a_{11}\mu^2)_i$, $(a_{12}\alpha)_i$, and $a_{2,i}$ over the 256 LOO optimizations (Figures 10 and 11). A low variability (low standard deviation) indicates a robust GC fit and thus good extrapolation capabilities of the developed model. The GC group parameters $(a_{11}\mu^2)_i$ are narrowly distributed with only a few outliers indicated by the standard deviation of maximal 0.07. This narrow distribution leads to similar predictions for the validation runs, because the model is particularly sensitive to the dipole scaling parameter $a_{11}\mu^2$. In comparison, the variation in the parameters $(a_{12}\alpha)_i$ is higher than in the parameters $(a_{11}\mu^2)_i$. With a standard deviation of maximal 0.21, the distribution is still narrow with a few outliers. The parameters $a_{2,i}$ are also narrowly distributed with only a few outliers. However, variation in $a_{2,i}$ is less critical than in $(a_{11}\mu^2)_i$,

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

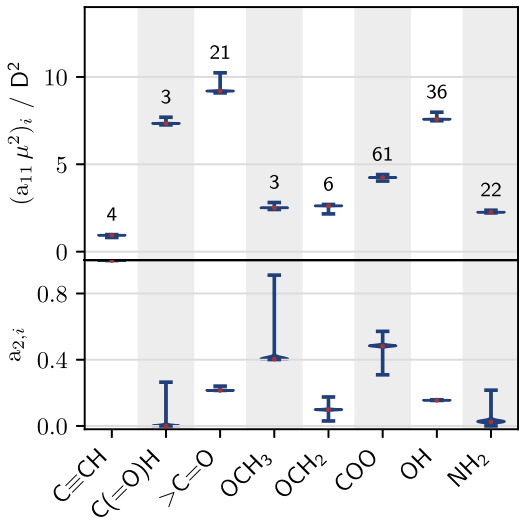


Figure 10. Distribution of optimal functional group parameter $(a_{11}\mu^2)_i$ and $a_{2,i}$ over 256 GC fits conducted for leave-one-out cross-validation. Red points indicate the value for the full GC fit. Numbers indicate in how many substances the respective functional group occurs.

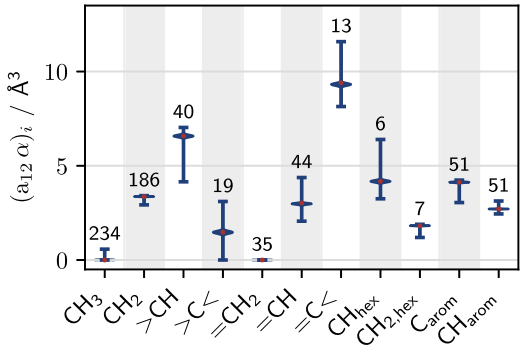


Figure 11. Distribution of optimal functional group parameter $(a_{12}\alpha)_i$ over 256 GC fits conducted for leave-one-out cross-validation. Red points indicate the values for the full GC fit. Numbers indicate in how many substances the respective functional group occurs.

because the permittivity model is less sensitive to the correlation integral parameter $a_{2,i}$.

Overall, accurate prediction results are obtained for nearly all substances, given the small set of adjustable parameters. The GC method is thus suitable for molecular design methods or for estimating the permittivity of substances without available experimental data.

5. CONCLUSION

This work presents a broadly applicable model for the calculation of temperature-dependent relative static permittivities based on perturbation theory. For this purpose, a previously proposed model for the relative static permittivity based on perturbation theory[41] is extended to nonpolar substances where the effective dipole is determined by polarizability effects. To establish a permittivity parameter database, the developed permittivity model is parametrized for 785 substances. The model extension enables accurate results, for both nonpolar and polar substances, over wide ranges of permittivities and temperatures.

To predict the model parameters based on the substance's molecular structure, we develop a group contribution method. The group contribution method accurately correlates the permittivities, despite the small set of parameters. The high predictive capabilities of the developed group contribution method are demonstrated by performing leave-one-out cross validation. For a straightforward application of the model, ready-to-use Python tools are provided for both the permittivity parameter database and the group contribution method.

The permittivity parameter database established in this work allows the accurate calculation of permittivities for many substances as required for high-throughput solvent screening. The group contribution method enables the calculation of permittivities of unknown substances from their molecular structure. Our permittivity model can be integrated into electrolyte equations of state, e.g., the electrolyte PC-SAFT equation of state.[11−15,63] The combination of an electrolyte equation of state with our permittivity model enables process design for various solvents, solvent screening, and solvent design for processes with electrolyte solutions.

■ ASSOCIATED CONTENT

Data Availability Statement

https://gitlab.ethz.ch/epse/molecular-design-public/paper-permittivity-based-on-perturbation-theory.git: Gitlab repository, contains all databases above and Python code to run the model developed in this work

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jced.3c00323.

Statistics of experimental data, further information on the permittivity parameter database (distribution of permanent dipole moments, liquid density correlation, selected permittivity-over-temperature diagrams, list of substances with higher errors), further information on the group contribution method (functional groups, substances included in the GC parametrization, comparison of weighting methods, selected permittivity-over-temperature diagrams) (PDF)

ParametersDensityCorrelation.csv: contains the parameters of the liquid density correlation and statistics on the model performance (cf., section 3.1.2); PermittivityParameterDatabase.csv: contains the parameters of the permittivity database and statistics on the model performance (cf., section 3.1); GCStatistics.csv: contains all substances included in the group contribution parametrization and statistics on the model performance (cf., section 3.2) (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

André Bardow − *Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, 8092 Zürich, Switzerland;* orcid.org/0000-0002-3831-0691; Email: abardow@ethz.ch

Joachim Gross − *Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, 70174 Stuttgart, Germany;* orcid.org/0000-0001-8632-357X; Email: gross@itt.uni-stuttgart.de

Authors

Lisa Rueben − *Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH*

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX

Zurich, 8092 Zürich, Switzerland; orcid.org/0000-0002-6914-4868

Johannes Schilling − *Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, 8092 Zürich, Switzerland;* orcid.org/0000-0001-8013-5439

Philipp Rehner − *Energy and Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, 8092 Zürich, Switzerland;* orcid.org/0000-0001-9750-5037

Simon Müller − *Institute of Thermal Separation Processes, Hamburg University of Technology, TUHH, 21073 Hamburg, Germany;* orcid.org/0000-0003-1684-6994

Timm Esper − *Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, 70174 Stuttgart, Germany;* orcid.org/0000-0002-2552-4391

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jced.3c00323

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Author J. Gross acknowledges funding by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075-390740016. Author P. Rehner acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 497566159. Open access funding provided by ETH Zurich.

■ REFERENCES

(1) Bülow, M.; Gerek Ince, N.; Hirohama, S.; Sadowski, G.; Held, C. Predicting Vapor−Liquid Equilibria for Sour-Gas Absorption in Aqueous Mixtures of Chemical and Physical Solvents or Ionic Liquids with ePC-SAFT. *Ind. Eng. Chem. Res.* **2021**, *60*, 6327−6336.

(2) Makeev, M. A.; Rajput, N. N. Computational screening of electrolyte materials: status quo and open problems. *Current Opinion in Chemical Engineering* **2019**, *23*, 58−69.

(3) Moura de Salles Pupo, M.; Kortlever, R. Electrolyte Effects on the Electrochemical Reduction of CO_2. *ChemPhysChem* **2019**, *20*, 2926−2935.

(4) Scheffczyk, J.; Redepenning, C.; Jens, C. M.; Winter, B.; Leonhard, K.; Marquardt, W.; Bardow, A. Massive, automated solvent screening for minimum energy demand in hybrid extraction−distillation using COSMO-RS. *Chem. Eng. Res. Des.* **2016**, *115*, 433−442.

(5) Scheffczyk, J.; Fleitmann, L.; Schwarz, A.; Lampe, M.; Bardow, A.; Leonhard, K. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chem. Eng. Sci.* **2017**, *159*, 84−92.

(6) Schilling, J.; Hopp, M.; Gross, J.; Bardow, A. Tailor-made solvents by integrated design of molecules and CO_2 absorption processes. *31st European Symposium on Computer Aided Process Engineering*; Elsevier, 2021; pp 197−202.

(7) Papadopoulos, A. I.; Badr, S.; Chremos, A.; Forte, E.; Zarogiannis, T.; Seferlis, P.; Papadokonstantakis, S.; Galindo, A.; Jackson, G.; Adjiman, C. S. Computer-aided molecular design and selection of CO_2 capture solvents based on thermodynamics, reactivity and sustainability. *Molecular Systems Design & Engineering* **2016**, *1*, 313−334.

(8) Gerlach, T.; Müller, S.; Smirnova, I. Development of a COSMO-RS based model for the calculation of phase equilibria in electrolyte systems. *AIChE J.* **2018**, *64*, 272−285.

(9) Müller, S.; González de Castilla, A.; Taeschler, C.; Klein, A.; Smirnova, I. Calculation of thermodynamic equilibria with the predictive electrolyte model COSMO-RS-ES: Improvements for low permittivity systems. *Fluid Phase Equilib.* **2020**, *506*, 112368.

(10) Kontogeorgis, G. M. *Thermodynamic Models for Industrial Applications*; Wiley, 2010; p 692.

(11) Cameretti, L. F.; Sadowski, G.; Mollerup, J. M. Modeling of Aqueous Electrolyte Solutions with Perturbed-Chain Statistical Associated Fluid Theory. *Ind. Eng. Chem. Res.* **2005**, *44*, 3355−3362.

(12) Held, C.; Cameretti, L. F.; Sadowski, G. Modeling aqueous electrolyte solutions. *Fluid Phase Equilib.* **2008**, *270*, 87−96.

(13) Held, C.; Reschke, T.; Mohammad, S.; Luza, A.; Sadowski, G. ePC-SAFT revised. *Chem. Eng. Res. Des.* **2014**, *92*, 2884−2897.

(14) Bülow, M.; Ascani, M.; Held, C. ePC-SAFT advanced - Part I: Physical meaning of including a concentration-dependent dielectric constant in the born term and in the Debye-Hückel theory. *Fluid Phase Equilib.* **2021**, *535*, 112967.

(15) Bülow, M.; Ascani, M.; Held, C. ePC-SAFT advanced − Part II: Application to Salt Solubility in Ionic and Organic Solvents and the Impact of Ion Pairing. *Fluid Phase Equilib.* **2021**, *537*, 112989.

(16) Kontogeorgis, G. M.; Maribo-Mogensen, B.; Thomsen, K. The Debye-Hückel theory and its importance in modeling electrolyte solutions. *Fluid Phase Equilib.* **2018**, *462*, 130−152.

(17) González de Castilla, A.; Müller, S.; Smirnova, I. On the analogy between the restricted primitive model and capacitor circuits. Part II: A generalized Gibbs-Duhem consistent extension of the Pitzer-Debye-Hückel term with corrections for low and variable relative permittivity. *J. Mol. Liq.* **2022**, *360*, 119398.

(18) Simon, H.-G.; Kistenmacher, H.; Prausnitz, J. M.; Vortmeyer, D. An equation of state for systems containing electrolytes and nonelectrolytes. *Chemical Engineering and Processing: Process Intensification* **1991**, *29*, 139−146.

(19) Fürst, W.; Renon, H. Representation of excess properties of electrolyte solutions using a new equation of state. *AIChE J.* **1993**, *39*, 335−343.

(20) Zuo, Y.-X.; Fürst, W. Prediction of vapor pressure for nonaqueous electrolyte solutions using an electrolyte equation of state. *Fluid Phase Equilib.* **1997**, *138*, 87−104.

(21) Uematsu, M.; Franck, E. U. Static Dielectric Constant of Water and Steam. *J. Phys. Chem. Ref. Data* **1980**, *9*, 1291−1306.

(22) Floriano, W. B.; Nascimento, M. A. C. Dielectric constant and density of water as a function of pressure at constant temperature. *Brazilian Journal of Physics* **2004**, *34*, 38−41.

(23) Kirkwood, J. G. The Dielectric Polarization of Polar Liquids. *J. Chem. Phys.* **1939**, *7*, 911−919.

(24) Fröhlich, H. Theory of dielectrics: Dielectric constant and dielectric loss. *Monographs on the physics and chemistry of materials*, 2nd ed.; Clarendon Press: Oxford, 1990; Vol. *42*.

(25) Harvey, A. H.; Prausnitz, J. M. Dielectric constants of fluid mixtures over a wide range of temperature and density. *J. Solution Chem.* **1987**, *16*, 857−869.

(26) Mohsen-Nia, M.; Amiri, H.; Jazi, B. Dielectric Constants of Water, Methanol, Ethanol, Butanol and Acetone: Measurement and Computational Study. *J. Solution Chem.* **2010**, *39*, 701−708.

(27) Kournopoulos, S.; Haslam, A. J.; Jackson, G.; Galindo, A.; Schoen, M. Molecular theory of the static dielectric constant of dipolar fluids. *J. Chem. Phys.* **2022**, *156*, 154111.

(28) Langenbach, K.; Kohns, M. Relative Permittivity of Dipolar Model Fluids from Molecular Simulation and from the Co-Oriented Fluid Functional Equation for Electrostatic Interactions. *Journal of Chemical & Engineering Data* **2020**, *65*, 980−986.

(29) Kohns, M. Molecular simulation study of dielectric constants of pure fluids and mixtures. *Fluid Phase Equilib.* **2020**, *506*, 112393.

(30) Maribo-Mogensen, B.; Kontogeorgis, G. M.; Thomsen, K. Modeling of dielectric properties of complex fluids with an equation of state. *journal of physical chemistry. B* **2013**, *117*, 3389−3397.

(31) Schweitzer, R. C.; Morris, J. B. The development of a quantitative structure property relationship (QSPR) for the prediction

of dielectric constants using neural networks. *Anal. Chim. Acta* **1999**, *384*, 285−303.
(32) Cocchi, M.; de Benedetti, P. G.; Seeber, R.; Tassi, L.; Ulrici, A. Development of Quantitative Structure−Property Relationships Using Calculated Descriptors for the Prediction of the Physicochemical Properties (n_D, ρ, bp, ε, η) of a Series of Organic Solvents. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1190−1203.
(33) Lee, A.; Kim, D.; Kim, K.-H.; Choi, S.-H.; Choi, K.; Jung, D. H. Elucidation of specific aspects of dielectric constants of conjugated organic compounds: a QSPR approach. *J. Mol. Model.* **2012**, *18*, 251−256.
(34) Sild, S.; Karelson, M. A general QSPR treatment for dielectric constants of organic compounds. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 360−367.
(35) Schweitzer; Morris. Improved quantitative structure property relationships for the prediction of dielectric constants for a set of diverse compounds by subsetting of the data set. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1253−1261.
(36) Achary, P. G. R. QSPR modelling of dielectric constants of π-conjugated organic compounds by means of the CORAL software. *SAR and QSAR in environmental research* **2014**, *25*, 507−526.
(37) Liu, J.-P.; Wilding, W. V.; Giles, N. F.; Rowley, R. L. A Quantitative Structure Property Relation Correlation of the Dielectric Constant for Organic Chemicals. *Journal of Chemical & Engineering Data* **2010**, *55*, 41−45.
(38) Gani, R. Group contribution-based property estimation methods: advances and perspectives. *Current Opinion in Chemical Engineering* **2019**, *23*, 184−196.
(39) Sheldon, T. J.; Adjiman, C. S.; Cordiner, J. L. Pure component properties from group contribution: Hydrogen-bond basicity, hydrogen-bond acidity, Hildebrand solubility parameter, macroscopic surface tension, dipole moment, refractive index and dielectric constant. *Fluid Phase Equilib.* **2005**, *231*, 27−37.
(40) Bouteloup, R.; Mathieu, D. Predicting dielectric constants of pure liquids: fragment-based Kirkwood-Fröhlich model applicable over a wide range of polarity. *Phys. Chem. Chem. Phys.* **2019**, *21*, 11043−11057.
(41) Neumaier, L.; Schilling, J.; Bardow, A.; Gross, J. Dielectric constant of mixed solvents based on perturbation theory. *Fluid Phase Equilib.* **2022**, *555*, 113346.
(42) Tani, A.; Henderson, D.; Barker, J. A.; Hecht, C. E. Application of perturbation theory to the calculation of the dielectric constant of a dipolar hard sphere fluid. *Mol. Phys.* **1983**, *48*, 863−869.
(43) Clausius, R. *Die Mechanische Behandlung der Electricität*; Vieweg +Teubner Verlag, 1879; p 368.
(44) Mossotti, O. F. Sull'influenza che l'azione di un mezzo dielettrico ha sulla distribuzione dell'elettricità alla superfice di più corpi elettrici disseminati in esso. *Memorie di matematica e di fisica della Società italiana delle scienze* **1850**, *24*, 49−74.
(45) Wertheim, M. S. Equilibrium Statistical Mechanics of Polar Fluids. *Annu. Rev. Phys. Chem.* **1979**, *30*, 471−501.
(46) Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids*; Oxford University Press, 1984.
(47) Huber, P. J. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* **1964**, *35*, 73−101.
(48) Hurvich, C. M.; Tsai, C.-L. A corrected Akaike Information Criterion for vector autoregressive model selection. *Journal of Time Series Analysis* **1993**, *14*, 271−279.
(49) Burnham, K. P.; Anderson, D. *Model Selection and Multi-Model Inference*; Springer, 2003; p 496.
(50) Thomson, G. H. The DIPPR databases. *Int. J. Thermophys.* **1996**, *17*, 223−232.
(51) Neese, F. The ORCA program system. *WIREs Computational Molecular Science* **2012**, *2*, 73−78.
(52) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Computational Molecular Science* **2018**, *8*, e1327.
(53) Yaws, C.; Pike, R. *Thermophysical Properties of Chemicals and Hydrocarbons*; Elsevier, 2009; pp 106−197.
(54) Bell, C.; Cortes-Pena, Y. R.; Contributors. Chemicals: Chemical properties component of Chemical Engineering Design Library (ChEDL), 2016−2021. https://github.com/CalebBell/chemicals.
(55) Dortmund Data Bank, 2021. www.ddbst.com.
(56) Frenkel, M.; Chirico, R. D.; Diky, V. V.; Dong, Q.; Frenkel, S.; Franchois, P. R.; Embry, D. L.; Teague, T. L.; Marsh, K. N.; Wilhoit, R. C. ThermoML An XML-Based Approach for Storage and Exchange of Experimental and Critically Evaluated Thermophysical and Thermochemical Property Data. 1. Experimental Data. *Journal of Chemical & Engineering Data* **2003**, *48*, 2−13.
(57) Frenkel, M.; Chiroco, R. D.; Diky, V.; Dong, Q.; Marsh, K. N.; Dymond, J. H.; Wakeham, W. A.; Stein, S. E.; Königsberger, E.; Goodwin, A. R. H. XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML) (IUPAC Recommendations 2006). *Pure Appl. Chem.* **2006**, *78*, 541−612.
(58) Gross, J.; Sadowski, G. Application of the Perturbed-Chain SAFT Equation of State to Associating Systems. *Ind. Eng. Chem. Res.* **2002**, *41*, 5510−5515.
(59) Sauer, E.; Stavrou, M.; Gross, J. Comparison between a Homo- and a Heterosegmented Group Contribution Approach Based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State. *Ind. Eng. Chem. Res.* **2014**, *53*, 14854−14864.
(60) Minkin, V. I., Osipov, O. A., Zhdanov, Y. A., Vaughan, W. E., Eds. *Dipole Moments in Organic Chemistry*; Springer US: Boston, MA, 1970.
(61) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *The properties of gases and liquids*, 5th ed.; McGraw-Hill: New York, 2001.
(62) Esper, T.; Bauer, G.; Rehner, P.; Gross, J. PCP-SAFT parameters of pure substances using large experimental databases. Submitted, **2023**.
(63) Bülow, M.; Ji, X.; Held, C. Incorporating a concentration-dependent dielectric constant into ePC-SAFT. An application to binary mixtures containing ionic liquids. *Fluid Phase Equilib.* **2019**, *492*, 26−33.

https://doi.org/10.1021/acs.jced.3c00323
J. Chem. Eng. Data XXXX, XXX, XXX−XXX